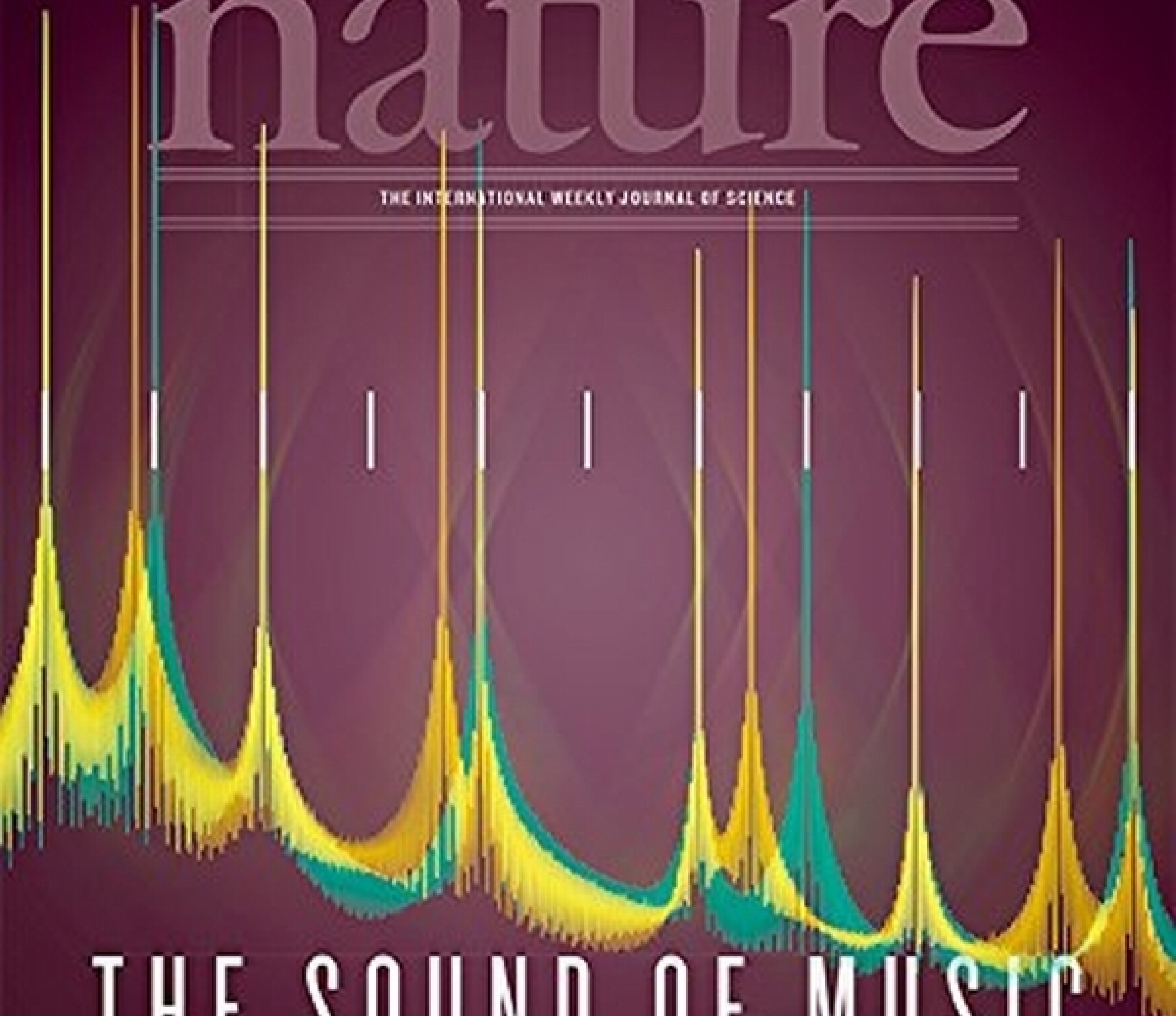
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



THE SOUND OF MUSIC

Studies of Bolivia's isolated Tsimane' society suggest a cultural origin for a preference for musical harmony PAGES 496 & 547

BIOTECHNOLOGY

STRANGE BREW

Meet the beer geeks crafting the perfect yeast

PAGE 484

POLITICS

LESSONS FROM BREXIT

Whither research in an age of demagogues?

PAGE 487

SCIENCE FICTION

THE MAKING OF FRANKENSTEIN

Science and society collide in Mary Shelley's epic

PAGE 490

NATUREASIA.COM

28 July 2016

Vol. 535, No. 7613

THIS WEEK

EDITORIALS

Prove the worth of basic research

European agencies are backing fundamental science and working to prove that it pays off. Other national and international bodies should follow suit.

The happy accidents that come from blue-skies research are gold dust for scientists, and help them to push back against political demands for applied work. Who doesn't know by now that we have basic research to thank for the World Wide Web? Who hasn't heard that curious researchers trying to work out how bacteria biochemically tick stumbled on the CRISPR–Cas9 gene-editing techniques that have gone on to transform biotechnology?

Still, political support for a thriving fundamental research base cannot be taken for granted. So two unexpected — and quite different — moves announced this month are worth noting and celebrating.

On 15 July, the hard-nosed European Investment Bank, which lends with favourable terms to European Union member states to support EU policy objectives, gave a massive loan to Greece to start up an agency for basic research. This not only provides a much needed moral boost for Greece, which has had to live for years with the label 'credit-unworthy', but it also sends a crystal-clear message to politicians around the world on the clear importance of pure science to a secure economic future.

Then, at the biennial European Open Science Forum in Manchester, UK, on 26 July, European Research Council (ERC) president Jean-Pierre Bourguignon announced that the council will start to monitor the outcomes of the research it funds. The ERC, which was founded in 2007 and awards sought-after grants that confer immense prestige on recipients, aims to systematically build a body of evidence to demonstrate the value of pure research beyond well-celebrated examples such as those mentioned above.

In the past 18 months, the ERC has quietly carried out a pilot effort to evaluate 199 of its first completed projects. It did not take the easy option of just looking at bibliometrics. It wisely took the more informative but more difficult option of asking experts not to get hung up on numbers, but to make judgements based on their expertise. They had to grade the scientific success of each project and assess its impact on the world outside science.

The results? The ERC seems to be a resounding success. (Although most of the reviewers had worked with the council before and so can't be classed as wholly independent.) Almost three-quarters of the projects were judged to have generated a scientific breakthrough or major scientific advance, and one-quarter had — or might have in the future — an impact on the economy, society or policymaking. The exercise cost a mere €200,000 (US$220,000), a tiny fraction of the ERC budget.

This is a very small qualitative study that has some flaws (see page 477), and the results cannot be extrapolated to the 6,000 or more grants, worth €9.8 billion, that the ERC has so far paid out. But the evaluation process is itself under constant review and many of its flaws should be ironed out in future rounds.

The results of the pilot will surprise few scientists, given the well-honed and widely admired selection procedures of the ERC. But as the years go by, they will add up to a convincing portfolio to present to politicians, showing that ERC spending on basic research is not wasted — it usually leads to scientific success, which in turn often leads to positive outcomes for society.

This type of retrospective audit is rare. And it is perhaps surprising that national research agencies around the world do not do it. The DFG in Germany, for example, feels that its own selection processes are reliable enough not to require further proof of this type — but then, in Germany, basic research is unusually well protected from the vagaries of politics. The time may be ripe for a modest investment like the ERC's to be more widely applied.

"The struggle between politicians and fundamental researchers is eternal."

The struggle between politicians and fundamental researchers is eternal, and understandably so. In democracies, politicians have to demonstrate to their electorates every five years or so that they have presided over serial successes and have not thrown away taxpayers' money on self-indulgent frippery. The scientific community has to find ways to continually show them that it is producing some of the successes. The strong endorsement of basic research by the European Investment Bank is a useful card that can be widely played to this end. And the ERC's example is one to follow: gather evidence for the worth of evidence-based arguments. ■

Cures for all

US lawmakers should give drug firms the confidence to test cancer therapies in children.

A cancer diagnosis is a shock, but adults with the disease can take some comfort in the numerous treatments available to them — both through clinical trials and as drugs that are already on the market. Children cannot. Because they make up only 1% of US patients with cancer, children are a low priority for pharmaceutical companies that want to launch an effective drug quickly. The hassle of a paediatric clinical trial may not seem worth it until after the drug has proved to be safe and effective in adults. This process can take decades, leaving children with therapies that are sometimes almost obsolete.

To access therapies early, parents of these children can turn to compassionate-use programmes, in which companies give experimental drugs to people who are in desperate need. In the United States, firms that agree to provide medicines in this way will ask the Food and Drug Authority for emergency permission, which is almost always granted.

This system, although helpful for some, is rife with complications.

Patients and their families report difficulties in applying for such programmes, and say that they rarely receive responses. Companies that withhold a drug — because it is in short supply or not right for a patient — can find themselves on the receiving end of critical social-media campaigns highlighting individual patients. And firms worry that if a person dies or is harmed while taking a drug, it could hurt the drug's chances of being approved. No one knows how many requests parents make and how often companies approve them, but anecdotally, firms often deny drugs on the grounds that they have not been tested in children.

Proper clinical trials for childhood cancer drugs are scarce. Designing a clinical trial is never simple, but adding children to the picture complicates the process immensely. Children are not just 'small adults' — they metabolize drugs in very different ways. It is difficult to predict from adult or animal studies whether a chemotherapy drug will be more or less toxic in a child, and at what dose. The process of obtaining informed consent for children participating in a trial can also be more complicated. And companies fear that the death of a child — even if unrelated to the treatment — could bring bad publicity for a new drug.

Recent years have seen attempts to make more drugs available to treat children. In the United States, a 2003 law known as the Pediatric Research Equity Act (PREA) requires that companies develop a plan for how they will test experimental drugs in children, although many trials are exempted. A second law, called the Best Pharmaceuticals for Children Act, motivates companies to perform paediatric clinical trials by granting an extra six months of market exclusivity for the adult drug.

Overall, these laws have been successful, leading to hundreds of drug labels being updated with information for use in children. But legal loopholes often prevent children with cancer from accessing new drugs. For instance, therapies for conditions that do not affect children — such as Alzheimer's disease — are exempt from the PREA. And exemptions intended for such diseases have been broadly applied to cancer. For example, therapies that are being trialled in adults with breast cancer are exempted because children do not get that cancer, even if the drug could treat a childhood cancer in a different organ.

Also exempted are drugs for 'orphan' diseases that affect fewer than 200,000 people in the United States. The number of orphan designations has skyrocketed in recent years — the improved ability to define the molecular basis of an individual's cancer means that diagnoses have become increasingly subdivided, and the majority of approved cancer drugs now carry this orphan designation.

"Legal loopholes often prevent children with cancer from accessing new drugs."

Legislation is now attempting to close those loopholes. The Research to Accelerate Cures and Equity (RACE) for Children Act, introduced to the US Congress on 14 July, would require companies to apply the PREA to any therapy with a molecular target that is relevant to both an adult and a childhood disease. It would also end the exemption for orphan diseases. Last July, the European Medicines Agency passed similar rules to make it more difficult for companies to avoid testing drugs in children. This applies when the disease has a common mechanism in adults and children, unless the drug is likely to be unsafe in children.

With Congress now out of session and focused on the upcoming US election, the RACE for Children Act is unlikely to advance before next year. But when lawmakers pick it up, they should also address problems with compassionate-use programmes — and ensure a transparent and useful process for people to gain access to unapproved drugs. They should also encourage companies to make more drugs available through market incentives, and provide increased protection should something go wrong. ■

On impact

Nature and the Nature journals are diversifying their presentation of performance indicators.

Metrics are intrinsically reductive and, as such, can be dangerous. Relying on them as a yardstick of performance, rather than as a pointer to underlying achievements and challenges, usually leads to pathological behaviour. The journal impact factor is just such a metric.

During a talk just over a decade ago, its co-creator, Eugene Garfield, compared his invention to nuclear energy. "I expected it to be used constructively while recognizing that in the wrong hands it might be abused," he said. "It did not occur to me that 'impact' would one day become so controversial."

As readers of *Nature* probably know, journal impact factors measure the average number of citations, per published article, for papers published over a two-year period. Journals do not calculate their impact factor directly — it is calculated and published by Thomson Reuters.

Publishers have long celebrated strong impact factors. It is, after all, one of the measures of their output's significance — as far as it goes.

But the impact factor is crude and also misleading. It effectively undervalues papers in disciplines that are slow-burning or have lower characteristic citation rates. Being an arithmetic mean, it gives disproportionate significance to a few very highly cited papers, and it falsely implies that papers with only a few citations are relatively unimportant.

These shortcomings are well known, but that has not prevented scientists, funders and universities from overly relying on impact factors, or publishers (*Nature*'s included, in the past) from excessively promoting them. As a result, researchers use the impact factor to help them decide which journals to submit to — to an extent that is undermining good science. The resulting pressures and disappointments are nothing but demoralizing, and in badly run labs can encourage sloppy research that, for example, fails to test assumptions thoroughly or to take all the data into account before submitting big claims.

The most pernicious aspect of this culture, as *Nature* has pointed out in the past, has been a practice of using journal impact factors as a basis for assessment of individual researchers' achievements. For example, when compiling a shortlist from several hundred job applicants, how easy it is to rule out anyone without a high-impact-factor journal in their CV.

How to militate against such a metrics-obsessed culture?

First, an approach that some have applied in the past and whose time has surely come. Applicants for any job, promotion or funding should be asked to include a short summary of what they consider their achievements to be, rather than just to list their publications. This may sound simplistic, but some who have tried it find that it properly focuses attention on the candidate rather than on journals.

Second, journals need to be more diverse in how they display their performance. Accordingly, *Nature* has updated its online journal metrics page to include an array of additional bibliometric data.

As a part of this update, for *Nature*, the Nature journals and *Scientific Reports*, we have calculated the two-year median — the median number of citations that articles published in 2013 and 2014 received in 2015. The median is not subject to distortion by outliers. (The two-year median is lower than the two-year impact factor: 24, down from 38, for *Nature*, for example.) For details, see go.nature.com/2arq7om.

Providing these extra metrics will not address the problem mentioned above of the diversity in citation characteristics between disciplines. Nor will it make much of a dent in impact-factor obsessions. But we hope that it will at least provide a better means of assessing our output, and put the impact factor in a better perspective.

However, whether you are assessing journals or researchers, nothing beats reading the papers and forming your own opinion. ■

WORLD VIEW *A personal take on events*

BRITISH COUNCIL

Resilient British science will withstand Brexit

UK scientists have had to fight to stay international before, and they must not stop now, says **Ehsan Masood***.*

There was no British delegate at the first meeting of the council to establish CERN, Europe's particle-physics lab, in May 1952. At the time, and in the earlier years when the project was being planned and launched, Britain's official position was that European cooperation offered little benefit to British physics, and the country preferred to go it alone.

Sound familiar? One month on from the 23 June vote for the United Kingdom to leave the European Union, the shock of a result that few in the research world expected shows no sign of easing.

The prevailing climate is uncertainty. Universities report panic among their staff. Despite reassurances from both the British government and the European Commission, the mere prospect of Brexit has already had a chilling effect on existing and future EU collaborations. One senior manager at a leading university told me how a long-standing professor from an EU country had asked about the risk of deportation.

We must not despair: in some respects, we have been here before. British centre-right governments have a tendency to view international cooperation as less important than competitiveness. Not only did one reject CERN, but a later Conservative government, led by Edward Heath, opposed the creation of the UN Environment Programme on the grounds that Britain already had the world's finest environment policies and that further action could harm industry. And, although it was a founding member of the UN Educational, Scientific and Cultural Organization, the United Kingdom pulled out in 1985 under the government of Margaret Thatcher. (It later rejoined.)

DESPITE UNCERTAINTY, SCIENCE MUST ASSUME THAT BREXIT WILL BE CARRIED OUT.

Throughout those years, Britain's scientists were no less outward-looking, no less collegial and no less innovative than ever. They just had to fight hard to retain their international links, as their successors and future generations will have to do.

This will take preparation and political engagement. Despite uncertainty, science must assume that the decision will be carried out. "Brexit means Brexit", Britain's incoming Prime Minister Theresa May has said repeatedly. And she has put key members of the campaign for Britain to leave the EU in ministerial positions where they will be negotiating exit.

Some in Britain's scientific community believe that there could be an exemption so that EU colleagues can continue to travel, live and work freely in the United Kingdom, but that is unlikely. The Brexit campaigners won on the basis of a promise that net migration to the United Kingdom, currently around 1 million every 3 years, will be cut to tens of thousands annually. That means that UK universities and scientists must prepare for movement within Europe to become more difficult.

Scientists are correct to warn about the loss of EU partnerships — one pro-vice-chancellor for research said that the greatest loss would be relationships, rather than money — but that will cut little ice with the government's Brexit-negotiating team. The entire argument to leave was built on the idea that the United Kingdom is too involved with Europe, and must therefore cut ties. In Parliament on 19 July, one of the leaders of the Brexit campaign, Michael Gove, said that British universities need not fear Brexit: they did not become among the best in the world by being part of any bureaucratic system.

It is crucial that science has a voice in the Brexit negotiations. The community must work through those of its researchers already in responsible roles in government departments, who can communicate with the new political leadership. If they don't, then the fringe voices of climate-change deniers and others still popular with many on the extremes of the Conservative Party will gain influence.

The UK Royal Society's approach must also evolve. Over the past two decades, the post of the society's foreign secretary has been a plum appointment. From a distance, it looked a bit like a colonial-era governorship, with nothing more onerous expected than to travel the world giving nice speeches and doing good. Now, the post's holder needs to start doing more to represent the interests of UK science at home and abroad, ready for when Brexit negotiations begin. The CERN example is pertinent: despite official opposition, observers from the Royal Society attended subsequent council meetings. This helped to smooth admission when the national mood changed.

Although UK researchers fear the loss of EU funds, including grants from the Horizon 2020 research-funding programme, ministers will probably agree on some kind of compensation: perhaps a global fund made open to scientists around the world, with grants for excellence and a spread of international partners. Big philanthropic funders, notably the Wellcome Trust, could also step in with funds, as the trust did with its contribution of a few hundred million pounds to the Joint Infrastructure Fund during the late 1990s, a previous time of austerity.

As important as it is to worry about loss of funds and free movement, the scientific community must take a more fundamental concern from the Brexit decision. The vote has exposed fears and divisions in British society, and a sizeable section of the political class exploited them. Scientists, universities and funders must do more to understand what caused these divisions, and at the same time help in the urgent task that is to heal them. ■

Ehsan Masood *is editor of* Research Fortnight. *His latest book is* The Great Invention: The Story of GDP and the Making and Unmaking of the Modern World.
e-mail: ehsan.masood@researchresearch.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

EVOLUTION

Early life liked it hot

The last universal common ancestor of all life — a microbe dubbed LUCA that existed around 3.5 billion years ago — probably resided in a hydrothermal vent that had low oxygen levels.

To find out how the organism lived, William Martin and his colleagues at Heinrich Heine University Düsseldorf in Germany reconstructed the evolutionary trees of more than 6 million genes from bacteria and archaea. They identified 355 protein families that were probably in LUCA's genome — these are involved in anaerobic metabolism and fixing carbon dioxide and nitrogen. This suggests that LUCA lived in an environment that was rich in hydrogen, CO_2 and iron, such as a hydrothermal vent.

LUCA may have depended heavily on the geochemistry of the vent to survive.

Nature Microbiol. **http://doi.org/bm2s (2016)**

CLAIRE N. SPOTTISWOODE

BEHAVIOURAL ECOLOGY

Bird helps people to find honey

A bird species responds to the specialized calls of human honey hunters, then leads them to bees' nests.

The greater honeyguide (*Indicator indicator*, **pictured** with honey hunter) benefits by eating the beeswax left behind by hunters after they break open bees' nests to get the honey. Claire Spottiswoode at the University of Cambridge, UK, and her colleagues found that 75% of nests located by honey hunters in Niassa National Reserve, Mozambique, were found with the help of the bird. Playing recordings of a traditional trill-grunt call made by the local Yao people while hunting doubled the chances of attracting a honeyguide to 66%, compared with recordings of other human or animal sounds. Overall, the call tripled the chances of finding a bees' nest to 54%.

The ability to understand and respond to human calls is not restricted to domesticated animals, the authors say.

Science 353, **387–389 (2016)**

CHEMISTRY

'Molecules' made from superatoms

Chemists have built structures similar to molecules but made up of superatoms — clusters of atoms with some of the same properties as atoms.

To make superatom 'molecules', Xavier Roy and his colleagues at Columbia University in New York City created cobalt–selenium clusters. They then attached two or three clusters to one another with 'arms' comprising various elements.

The team hopes that the technique can be used to make materials with tailored properties such as heat conduction or the ability to store information magnetically.

Nano Lett. **http://doi.org/bmw5 (2016)**

PLANETARY SCIENCE

Revived telescope finds 104 planets

Astronomers have spied 104 new worlds in the Milky Way using NASA's Kepler Space Telescope.

Part of Kepler broke down in 2013, but engineers managed to repair it and send it on a fresh mission, dubbed K2. This latest discovery, from Ian Crossfield at the University of Arizona in Tucson and an international team, is the biggest so far for the K2 mission. The team reports numerous planetary candidates, and confirmed more than 100 as exoplanets using additional observations from ground-based telescopes. They found that the majority of planets are smaller than Neptune, and probably have thick atmospheres and rocky cores. Nearly 40 have a radius that is about twice that of Earth or smaller, and 4 of those orbit the same red dwarf star.

Two of the planets orbiting the red dwarf could have irradiation levels that are similar to Earth's, making life on those planets a possibility, according to the authors.

Astrophys. J. Suppl. Ser. **(in the press); preprint at http://arxiv.org/abs/1607.05263 (2016)**

NEURODEGENERATION

How immune cells clear amyloid

Three key proteins allow immune cells in the brain to clear out a protein called

DANIEL J COX/GETTY

amyloid-β, which is a hallmark of Alzheimer's disease.

The immune cells, called microglia, normally absorb and digest amyloid-β after lipoproteins called APOE and CLU attach to it. Lino Gonzalez and Morgan Sheng of drug firm Genentech in South San Francisco, California, and their colleagues looked for interactions between human proteins, and found that a receptor on microglia called TREM2 binds to APOE and CLU. Mouse microglia lacking Trem2 were less effective at absorbing amyloid–lipoprotein complexes, and digested them more slowly than normal cells did. Microglia taken from people carrying a TREM2 mutation were also less able to take up the complexes.

Mutations in the genes that encode TREM2, APOE and CLU have been linked to Alzheimer's, and so the results show how these different genetic risk factors could be linked.

Neuron **91, 328–340 (2016)**

PALAEOCLIMATE

Analysing ancient air in salt crystals

Gas trapped in ancient bubbles reveals that Earth's atmosphere was rich in oxygen up to 200 million years earlier than models have predicted, well before animal diversity exploded.

Bubbles in salt crystals called halites (**pictured**) that formed millions of years ago can provide clues to ancient climates. Nigel Blamey of Brock University in St Catharines, Ontario, Canada, and his colleagues studied pockets of air inside 815-million-year-old halites from southwest Australia. The air contained nearly 11% oxygen, more than expected for that time period.

DIDIER DESCOUENS/CC BY-SA 4.0

The authors suggest that high oxygen levels drove animal evolution, rather than the other way around.

Geology **http://doi.org/bmt3 (2016)**

OPTICS

Human eye sees single photons

People can perceive flashes of light as feeble as a single photon.

Alipasha Vaziri at the Rockefeller University in New York City and his colleagues asked three volunteers to stare into an optical system in the dark and listen to two sounds, one of which was sometimes accompanied by the emission of a photon. During more than 30,000 trials, the participants correctly identified a photon more frequently than would be expected if they had guessed at random.

Going forwards, the team plans to test how the human visual system responds to photons in various quantum states.

Nature Commun. **7, 12172 (2016)**

For more on this story, see go.nature.com/2anfrrf

MARINE ECOLOGY

Ice algae key to Arctic food web

Even creatures living many metres below the Arctic Ocean's surface rely on algae that grow in sea ice and so, like those living near the surface, may feel the negative effects of shrinking ice.

A team led by Doreen Kohlbach of the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research in Bremerhaven, Germany, collected small crustaceans and other animals at different depths in the central Arctic Ocean and analysed their fatty-acid content to identify the source of carbon in their diets. Organisms living at the water–ice interface (such as *Apherusa glacialis*; **pictured**) got as much as 92% of their carbon from the ice algae. But creatures sampled at up to 50 metres below the surface got 14–55% of their carbon from the algae.

Melting sea ice means a shrinking habitat for algae, which could lead to decreased nourishment for the entire Arctic food web.

Limnol. Oceanogr. **http://doi.org/bmtq (2016)**

NEUROSCIENCE

Neurons compete to make memories

Memories that are formed within a few hours of one another can be encoded by a shared set of neurons.

Sheena Josselyn and Paul Frankland at the Hospital for Sick Children in Toronto, Canada, and their colleagues exposed mice to two fear-inducing experiences, each consisting of a different musical tone paired with a mild electric shock to the foot. When the events occurred within 6 hours of each other, the associated memories were encoded by some of the same neurons in the lateral amygdala — a brain area involved in fear memory. However, fear memories separated by 24 hours were encoded by separate groups of neurons.

The researchers found that neurons stimulated by the first experience transiently suppressed the excitability of neighbouring neurons, increasing the chances that the same cells will respond to a second experience that follows soon afterwards.

Science **353, 383–387 (2016)**

ENERGY

Carbon capture makes electricity

A battery developed from widely available materials produces electricity by capturing carbon dioxide from gas mixtures.

Wajdi Al Sadat and Lynden Archer at Cornell University in Ithaca, New York, designed an aluminium-based electrochemical cell. At the cell's cathode, oxygen is reduced to form a superoxide, which then binds CO_2 and combines with aluminium from the anode to form aluminium oxalate.

The authors estimate that for each kilogram of aluminium, more than 9 kilograms of CO_2 can be captured from flue gas and transformed to generate 3.6 kilowatt-hours of electricity. This offers a strategy to reduce CO_2 emissions while producing power.

Sci. Adv. **2, e1600968 (2016)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

KARIM SAHIB/AFP/GETTY

Solar plane achieves historic global flight

The world's first fuel-free flight around the world, in a solar-powered aeroplane, concluded successfully on 26 July, when the Solar Impulse 2 craft landed in Abu Dhabi. Pilots Bertrand Piccard and André Borschberg flew 40,000 kilometres in 17 separate flights over 16 months. Among the records set during the circumnavigation was Borschberg's 118-hour stretch of solo endurance flying between Nagoya, Japan, and Hawaii. Solar Impulse took off from Cairo on 24 July for the final leg, and endured turbulence over the Saudi desert before touching down with Piccard at the controls. The craft has 17,000 solar cells and a wingspan of 72 metres.

EVENTS

Turkey in turmoil

The clampdown on political opposition by Turkish President Recep Tayyip Erdoğan in the wake of the failed military coup on 15 July spread to the education sector last week. The Turkish Council of Higher Education (YÖK) told all 1,577 of the country's university deans to resign their posts, and placed a temporary international-travel ban on Turkish scholars (see *Nature* http://doi.org/bm4d; 2016). Erdoğan has also closed 15 universities that he alleges are closely tied to the Gülen religious movement, which he considers to be behind the coup. Several academic organizations in Europe, including the German Rectors' Conference and the European Federation of Academies of Science and Humanities, have condemned the attack on academic independence.

SciDev.Net cuts

SciDev.Net, a science and technology news service for the developing world, announced on 22 July that it is shedding about 80% of its London-based staff, including much of the editorial team, owing to a lack of funding. SciDev.Net was launched in 2001 with funding from the international development agencies of Britain, Sweden and Canada, and with guidance from journals including *Nature*. It said that its decision to scale down its London operations reflected a commitment to preserve its regional services.

EU collaboration

University leaders from across Europe published a joint statement on 22 July calling for continent-wide collaboration to continue in the wake of the United Kingdom's vote to leave the European Union last month. The statement, signed by representatives from 24 national and international university consortia, calls for the exchange of ideas and people, as well as the pooling of infrastructure, data and expertise, to continue as part of the United Kingdom's new relationship with the rest of Europe.

Digital free speech

The Electronic Frontier Foundation (EFF), a group that fights for civil liberties in the digital space, has sued the US government over a law that it says restricts researchers' ability to conduct IT science. It says that parts of the Digital Millennium Copyright Act that restrict access to computer code violate the first amendment to the US Constitution, which guarantees freedom of speech. In particular, the law "stymies creators, academics, inventors, and researchers", says the EFF, which filed the lawsuit on 21 July. Among those bringing the suit is Matthew Green, an IT-security researcher at Johns Hopkins University in Baltimore, Maryland, who says that the law blocks the work of scientists such as him who want to probe the security of digital systems.

Olive-tree deaths

The European Commission sent Italy a stern warning on 22 July, telling it to fully implement plans to stop the bacterium *Xylella fastidiosa* — currently devastating olive groves in the Puglia region — from spreading to other parts of Italy and Europe. In May, the commission and the Italian government agreed a containment plan that includes controversial measures such as felling infected trees and nearby healthy ones, but the commission says that the plan is not being put fully into practice. It argues that *X. fastidiosa* is one of the world's most dangerous plant bacteria, and its spread would have a huge economic impact.

RESEARCH

Still in the dark

The most sensitive experiment yet designed to catch particles of dark matter particles has concluded its

search empty handed. After a 20-month run, the Large Underground Xenon (LUX) experiment in South Dakota announced on 21 July that it had seen no collisions of 'weakly interacting massive particles' (WIMPs) with its 370 kilograms of liquid xenon. The researchers say that WIMPs could still exist, but might interact more rarely than current experiments can detect. Preparations are under way to build a much larger dark-matter detector, called LZ.

Hottest year

This year is on track to beat 2015 as the warmest year since systematic measurements began. The first 6 months of the year were 0.2 °C warmer than the first half of 2015, the World Meteorological Organization said on 21 July. The average global temperature was 1.3 °C above pre-industrial levels, and 1.05 °C above the twentieth-century average, according to data sets maintained by NASA, the US National Oceanic and Atmospheric Administration and the UK Met Office. A strong 2015–16 El Niño warming event in the Pacific Ocean contributed to the record temperatures, which are driven by rising atmospheric concentrations of greenhouse gases.

MARY ALTAFFER/AP PHOTO

PEOPLE

Clinton picks Kaine

Hillary Clinton, the presumptive Democratic nominee for US president, named Virginia senator Tim Kaine (**pictured**) as her vice-presidential running mate on 22 July. Kaine previously served as governor of Virginia, which has been among the key battleground states in recent elections. He advocated for action on climate change as governor and had worked on education and foreign policy in the Senate.

Agriculture chief

French President François Hollande ignored scientists' furious protests about his choice to head the French National Institute for Agricultural Research (INRA) for the next four years, and allowed his cabinet to confirm the appointment of Philippe Mauguin on 19 July. Mauguin, a former chief of staff to the agriculture minister, takes over "in a climate of strong mistrust", says the INRA branch of the trade union CGT. It fears that Mauguin will focus on innovation at the expense of basic research, which would be potentially "deadly" for the agency's position as a leader in agricultural research.

POLICY

Porpoise protection

Environmental groups have welcomed the latest attempts by Mexico's National Commission of Aquaculture and Fishing (Conapesca) to save the vaquita porpoise (*Phocoena sinus*), the world's most endangered marine mammal. On 19 July, Conapesca announced rules banning 'gill nets' and night fishing in the northern part of the Gulf of California, where the animals live. Only 60 vaquita were thought to have been alive in the gulf in December 2015, down from 97 in 2014; gill nets are blamed for much of the decline.

Aircraft emissions

The US Environmental Protection Agency (EPA) paved the way for new climate regulations on 25 July by declaring that greenhouse-gas emissions from aircraft endanger public health and the environment. The finding comes five months after the United Nations' International Civil Aviation Organization (ICAO) proposed the first greenhouse-gas standards for aircraft. Environmentalists criticized the ICAO proposal as weak and are pressuring the EPA to suggest stronger regulations. The EPA said that its standards would be "at least as stringent" as those proposed by the ICAO.

COMING UP

27 JULY–10 AUGUST
Young scientists from around the world take part in the London International Youth Science Forum 2016, hosted by Imperial College London.
www.liysf.org.uk

1–12 AUGUST
Physicists discuss LIGO's gravitational-wave detection at the Kavli Institute for Theoretical Physics in Santa Barbara, California.
go.nature.com/2afhjwa

Predator plan

New Zealand hopes to be free of introduced predators — which include rats, stoats and possums — by 2050, according to a plan announced on 25 July. The government will invest an initial NZ$28 million (US$20 million) and work with the private sector to rid the country of the predators, which it says kill 25 million native birds each year. Conservation minister Maggie Barry says that the plan will need technology that does not yet exist.

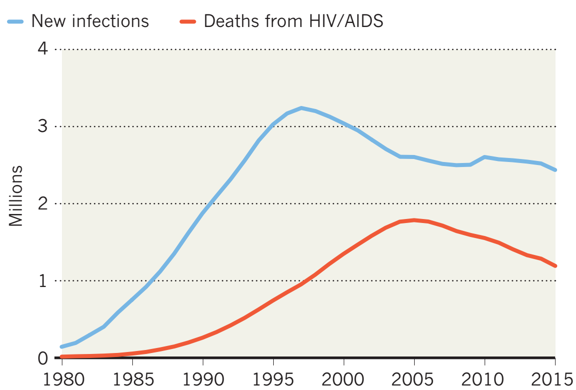
TREND WATCH

Deaths from HIV/AIDS have declined steadily around the world over the past decade — but the annual rate of new infections has stayed much the same, a global data analysis published on 19 July shows (H. Wang *et al. Lancet HIV* http://doi.org/bm2p; 2016). The number of new HIV infections peaked at 3.3 million in 1997, and dropped by an average of about 2.7% each year to around 2.5 million in 2005. But the infection rate has stagnated since then. In 74 countries, it has increased.

SOURCE: WANG *ET AL.* (2016)

DECLINE IN NEW HIV INFECTIONS STALLS

Although deaths from HIV/AIDS have dropped steadily over the past ten years, the global annual number of new infections has stayed largely the same.



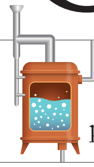
➲ NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

ASTRONOMY NASA seeks fresh chemical signatures of alien life **p.474**

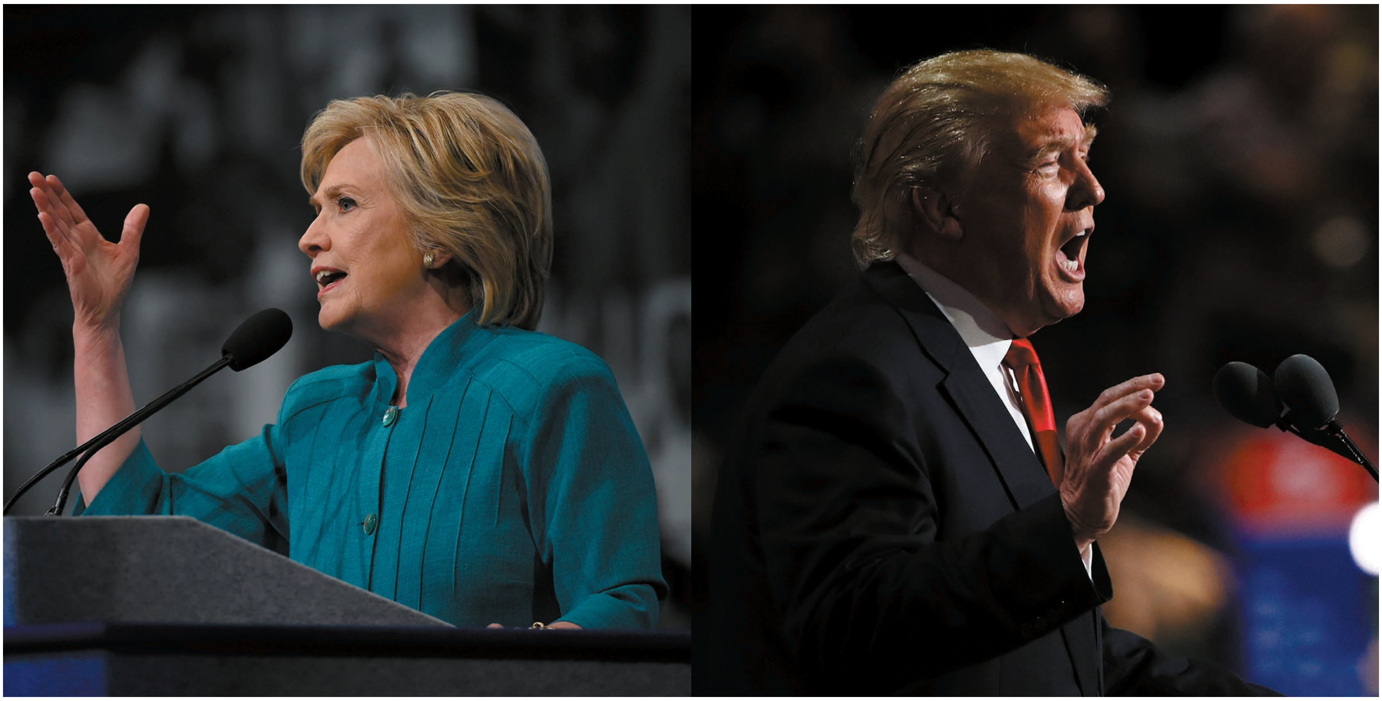
MEDICINE Chinese team approved to test CRISPR gene-editing in people **p.476**

SPACE Quantum satellite is step towards secure communications **p.478**



GENETICS The Belgian beer lab that is seeking the perfect yeast **p.484**

L: ETHAN MILLER/GETTY; R: WIN MCNAMEE/GETTY



Hillary Clinton and Donald Trump are battling for the White House.

US ELECTION

Trump vs Clinton: worlds apart on science

Presidential candidates begin to make clear their stark differences on climate change, energy production and stem-cell research.

BY JEFF TOLLEFSON

Science is slowly coming into focus in the US presidential campaign. Although neither Republican Donald Trump nor Democrat Hillary Clinton has emphasized core research issues, the candidates — and their parties — are beginning to flesh out their positions on climate change, education, biomedical research and other topics that involve the scientific community.

Trump's pick of Indiana Governor Mike Pence as his running mate on 15 July signalled a sharp turn towards the Republican party's conservative base. Pence, a self-described Christian conservative, has questioned the existence of climate change, waffled on evolution and criticized President Barack Obama for supporting embryonic-stem-cell research. His new role aligns with the hard-line policy platform adopted at the Republican convention, where Trump officially became the party's nominee on 19 July.

If Trump wins, Pence's rise could embolden conservative Republicans to seek new limits on federal funding for embryonic-stem-cell research. But predicting how Trump would govern is a dangerous parlour game, says Michael Werner, executive director of the Alliance for Regenerative Medicine, an advocacy group in Washington DC: "We really don't know what a Trump–Pence administration would do."

It's a common refrain. Deciphering Trump's views on core science issues has been difficult given the free-wheeling style of his populist campaign. He has often seemed to focus more on taunting the political establishment than on staking out policy positions. By contrast, the Clinton campaign has consulted dozens ▶

of scientists on topics that include health, education and the environment.

"Trump doesn't have a prominent policy shop and a prominent set of policy advisers," says Douglas Holtz-Eakin, who counselled Republican senator John McCain (Arizona) on economic policy during his failed 2008 presidential bid. "Clinton has a vast bureaucracy and a ten-point plan for going out to lunch, so they couldn't be more different."

The two candidates — whose campaign staff declined multiple interview requests — also seem to think very differently about the role of science. Although Clinton has described science and innovation as a foundation for the future, science funding seems to be an afterthought for Trump, says John Karsten, coordinator of the Center for Technology Innovation at the Brookings Institution, a think tank in Washington DC. Instead, the Republican has focused on issues such as national security, immigration and crumbling infrastructure.

Climate change is one of the few science topics that has grabbed the campaign spotlight — in part because of Republican anger over Obama's regulations to limit greenhouse-gas emissions from power plants, vehicles and oil and gas development. Clinton's climate and energy proposals would largely maintain the current course; by contrast, in a major policy speech on 26 May, Trump promised to roll back Obama's "totalitarian" regulations and withdraw the United States from the Paris climate agreement. Trump, who has long denied mainstream climate science, also said that his administration will focus on "real environmental challenges, not phony ones".

SPLIT TICKETS

This yawning philosophical divide is apparent in the party platforms that the Republicans and Democrats developed ahead of their nominating conventions this month. Environmentalists have criticized the Republican platform for labelling coal a "clean" energy source, even though it produces more carbon dioxide emissions per unit of energy than any other fossil fuel. Democrats, meanwhile, are poised to adopt a platform this week at their national convention that calls for using "every tool available to reduce emissions now".

> ***"Climate is going to be talked about in this campaign."***

"Climate is going to be talked about in this campaign, because the candidates have distinctly different positions," says Michael Oppenheimer, a climate scientist at Princeton University in New Jersey who is advising the Clinton team. Although his workload was light during primary season, Oppenheimer anticipates questions from the campaign about how global warming might affect certain regions, or the extent to which an extreme weather event might be related to global warming.

Some experts say that the Democratic party's adoption of science as a campaign issue — which Obama kick-started in 2008 — risks further polarizing thorny policy debates around scientific issues such as global warming. "The Democrats found that science was a good thing for them, just like historically strong support for the military was good for the Republicans," says Daniel Sarewitz, co-director of Arizona State University's Consortium for Science, Policy and Outcomes in Washington DC (and a regular contributor to *Nature*). "If the Democrats are the party of science, and you are a Republican, what does that make you think?"

But Holtz-Eakin says that the Trump campaign's apparent decision to forgo science advice is a reflection of Trump himself, not of Republican priorities. In 2008, he notes, the McCain campaign consulted scientists to formulate its positions on issues such as global warming — just as Clinton has done.

With just over three months until the election, there is still a chance that Trump will assemble his own coterie of science advisers, says Andrew Rosenberg, who heads the Center for Science and Democracy at the Union of Concerned Scientists in Cambridge, Massachusetts. Doing so not only informs policy positions, it builds relationships that are useful after the election, when the winning candidate begins to assemble a government.

"These things widen the network," Rosenberg says. "I know it's happening with the Clinton campaign, and at some point I would expect it would happen with the Trump campaign." ■

ASTRONOMY

How to hunt for alien life

Astrobiologists try to determine the chemical signature of life on other worlds.

BY ALEXANDRA WITZE

In the search for life beyond Earth, false alarms abound. Researchers have generally considered, and rejected, claims ranging from a 1970s report of life on Mars to the 1990s 'discovery' of fossilized space microbes in a meteorite.

Now, inspired by the detection of thousands of planets beyond the Solar System, NASA has started a fresh effort to learn how to recognize extraterrestrial life. The goal is to understand what gases alien life might produce — and how Earth-bound astronomers might detect such 'biosignatures' in light passing through the atmospheres of planets trillions of kilometres away (see 'Searching for alien life').

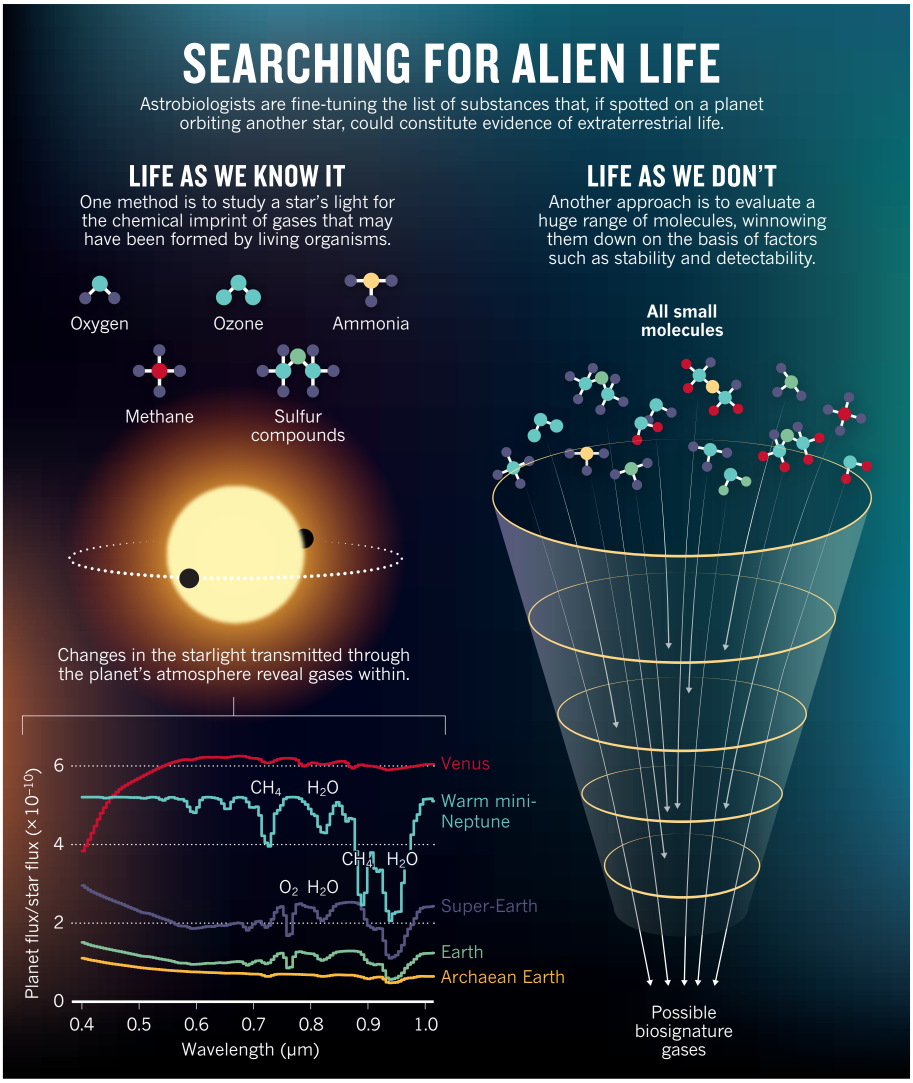
The agency will convene a workshop this week in Seattle, Washington, with the ultimate goal of advising a NASA exoplanet group on how to avoid embarrassing errors in the future. "We have to come together and determine what good evidence of life on another planet could be," says Shawn Domagal-Goldman, one of the workshop's organizers and an astronomer at NASA's Goddard Space Flight Center in Greenbelt, Maryland.

The exercise comes at a crucial time, as astronomers grapple with how to interpret exoplanet data from the next generation of telescopes. Some scientists are working to understand how nature could produce archetypal biosignature gases, such as oxygen, in the absence of living organisms. Others are trying to think as expansively as possible about the types of biochemistry that could sustain life.

"We could fool ourselves into thinking a lifeless planet has life — or we could be missing life because we don't really understand the context of what could be produced on another planet," says Sarah Rugheimer, an astronomer at the University of St Andrews, UK.

Detecting a biosignature gas is just the first step to understanding what could be happening on an exoplanet. Each world has its own combination of physical and chemical factors that may or may not lead to life, says Victoria Meadows, an astronomer at the University of Washington in Seattle. "Planets are hard, and we shouldn't think they are all going to be the same or reveal their secrets very easily," she says.

A planet's environment is key. Some Earth-sized planets orbit M dwarf stars — the most common type of star in the Galaxy — at the right distance to harbour liquid water. But Meadows' collaborators have shown[1] that photochemical reactions can send water into the planet's atmosphere and then break off its hydrogen, which escapes into space. What's left is a thick blanket of oxygen that might seem as



SOURCE: CHART: S. SEAGER & W. BAINS *SCI. ADV.* **1**, E1500047 (2015); CONE: REF. 5

if it came from living organisms, but results from a runaway greenhouse effect.

There are ways to tell. The runaway greenhouse would create an atmosphere thousands of times denser than Earth's, in which O_2 molecules collide to produce O_4. So spotting O_4 in a planet's atmosphere could be a clue that the oxygen does not, in fact, come from life, Meadows' team reported this year[2].

Another method is to draw up a list of alternative biosignature gases — things not as obvious as oxygen that might be made by organisms under certain conditions. These include dimethyl sulfide[3], which is produced by Earthly phytoplankton, or even ammonia[4]. On a cold alien planet, organisms might make the gas using the same chemical process as industrial manufacturers.

At the Massachusetts Institute of Technology in Cambridge, astronomer Sara Seager has begun to examine 14,000 compounds that are stable enough to exist in a planetary atmosphere. She and her colleagues are winnowing down their initial list of molecules using criteria such as whether there are geophysical ways to send the compound into the atmosphere[5].

"We're doing a triage process," says Seager. "We don't want to miss anything."

The Seattle meeting aims to compile a working list of biosignature gases and their chemical properties. The information will feed into how astronomers analyse data from NASA's James Webb Space Telescope, slated for launch in 2018. The telescope will be able to look at only a handful of habitable planets, but it will provide the first detailed glimpse of what gases surround which world, says Nikole Lewis, an astronomer at the Space Telescope Science Institute in Baltimore, Maryland.

No single gas is likely to be a slam-dunk indicator of alien life. But Domagal-Goldman hopes that the workshop will produce a framework for understanding where scientists could trip themselves up. "We don't want to have a great press release," he says, "and then a week later have egg on everybody's faces." ■

1. Luger, R. & Barnes, R. *Astrobiology* **15,** 119–143 (2015).
2. Schwieterman, E. W. *et al. Astrophys. J.* **819,** L13 (2016).
3. Domagal-Goldman, S. D., Meadows, V. S., Claire, M. W. & Kasting, J. F. *Astrobiology* **11,** 419–441 (2011).
4. Seager, S., Bains, W. & Hu, R. *Astrophys. J.* **775,** 104 (2013).
5. Seager, S., Bains, W. & Petkowski, J. J. *Astrobiology* **16,** 465–485 (2016).

EPIDEMIOLOGY

Brazil's birth-defects puzzle

Zika virus might not be only factor in reported microcephaly surge.

BY DECLAN BUTLER

Government researchers in Brazil are set to explore the country's peculiar distribution of Zika-linked microcephaly — babies born with abnormally small heads.

Zika virus has spread throughout Brazil, but extremely high rates of microcephaly have been reported only in the country's northeast. Although evidence suggests that Zika can cause microcephaly, the clustering pattern hints that other environmental, socio-economic or biological factors could be at play.

"We suspect that something more than Zika virus is causing the high intensity and severity of cases," says Fatima Marinho, director of information and health analysis at Brazil's ministry of health. If that turns out to be true, it could change researchers' assessment of the risk that Zika poses to pregnant women and their children.

The idea has long been on Brazilian researchers' radar, but the enquiry marks the first time that scientists at the health ministry have taken up the hypothesis. The ministry has asked Oliver Brady, an epidemiologist at the London School of Hygiene & Tropical Medicine, ▶

▶ and Simon Hay, director of geospatial science at the Institute for Health Metrics and Evaluation in Seattle, Washington, to collaborate with researchers in Brazil. "The aim is to understand why we are only observing elevated rates in the northeast," says Brady, who flew into Brasilia this month to begin work.

The northeast was where the first reported surge in microcephaly cases in Brazil began a year ago. Health officials had expected that they would later see the same high rates in other parts of the country. "We were expecting an explosion of birth defects," says Marinho.

But as of 20 July, almost 90% of the 1,709 confirmed cases of congenital microcephaly or birth defects of the central nervous system reported in Brazil since last November were in a relatively small area: in the coastal hinterland of the country's northeastern tip. Particularly surprising, says Marinho, is that just three cases have been confirmed in Brazil's second-most populous state, Minas Gerais, which borders the most-affected part of the northeast region. Poor data on the scale and timing of Zika outbreaks across Brazil make it hard to tell whether increases in microcephaly elsewhere might have been delayed — but ministry scientists now think that the northeast represents a marked outlier, she says.

FELIPE DANA/AP

A health worker sprays insecticide to combat the mosquito that spreads Zika.

There are many hypotheses about what might be going on. Marinho says that her team's data, submitted for publication, hint that socio-economic factors might be involved. For example, the majority of women who have had babies with microcephaly have been young, single, black, poor and tend to live in small cities or on the outskirts of big ones, she says.

Another idea is that co-infections of Zika and other viruses, such as dengue and chikungunya, might be interacting to cause the high intensity of birth defects in the area.

In a paper published last month, researchers from Brazilian labs noted a correlation between low vaccination rates for yellow fever and the microcephaly clusters (L. P. de Goes Cavalcanti *et al. J. Infect. Dev. Countries* **10,** 563–566; 2016). Because yellow fever and Zika are in the same virus family, the scientists speculate that the vaccine might provide some protection against Zika.

And the Brazilian doctor who was the first to report a firm link between Zika and microcephaly — Adriana Melo at IPESQ, a research institute in Campina Grande — has another idea. In a preprint posted on the bioRxiv server on 15 July, Melo and her colleagues at the Federal University of Rio de Janeiro reported finding bovine viral diarrhoea virus (BVDV) proteins in the brains of three fetuses with microcephaly (F. C. S. Nogueira *et al.* Preprint at bioRxiv http://doi.org/bm4c; 2016).

BVDV causes birth defects in cattle but is not known to infect people. Melo and her team suggest that Zika infection might make it easier for BVDV to cause infections; however, they haven't ruled out the possibility that their findings might be due to contamination.

The Brazilian health ministry's study will test for BVDV among other ideas, says Brady. Researchers will reanalyse raw data on microcephaly cases, and will model connections with possible cofactors such as socio-economic status, water contamination and mosquito-borne diseases. Most of this information comes from health-ministry databases, but the team will also study experimental data, such as how people's immune response may change after past infection with other viruses such as dengue.

Until more is known about Zika and the causes of increased microcephaly rates in Brazil's northeast, public-health actions and advice must err on the side of precaution, says Ian Lipkin, a virologist and outbreak specialist at Columbia University in New York City. ■

BIOMEDICINE

First trial of CRISPR in people

Chinese team approved to test gene-edited cells in people with lung cancer.

BY DAVID CYRANOSKI

Chinese scientists are on the verge of being first in the world to inject people with cells modified using the CRISPR–Cas9 gene-editing technique.

A team led by Lu You, an oncologist at Sichuan University's West China Hospital in Chengdu, received ethical approval to test the cells in people with lung cancer on 6 July, and plans to start the trial next month.

That timeline puts the proposal ahead of a planned US trial to test CRISPR–Cas9-modified cells, also for the treatment of cancer. "It's an exciting step forward," says Carl June, a clinical researcher in immunotherapy at the University of Pennsylvania in Philadelphia.

Last month, the US trial was approved by an advisory panel of the US National Institutes of Health (NIH) but had yet to receive a green light from the US Food and Drug Administration (FDA) and a university review board. There have also been a number of human clinical trials using an alternative gene-editing technique, including one led by June, that have helped patients to combat HIV — but none so far has used CRISPR.

The Chinese trial will enrol patients who have metastatic non-small cell lung cancer and for whom chemotherapy, radiation therapy and other treatments have failed. "This technique is of great promise in bringing benefits to patients," says Lu.

CHROMOSOME SNIP

Lu's team will extract immune cells called T cells from the participants' blood, and use CRISPR–Cas9 technology — which pairs a molecular guide able to identify specific genetic sequences on a chromosome with an enzyme that can snip the chromosome at that spot — to knock out a specific gene in the

cells. The target gene encodes a protein called PD-1 that normally acts as a check on the cell's capacity to launch an immune response.

The gene-edited cells will then be multiplied in the lab and re-introduced into the patient's bloodstream, where, the team hopes, they will home in on the cancer. The proposed US trial similarly involves knocking out the gene for PD-1, but also includes knocking out a second gene and inserting a third.

Last year, the FDA approved two antibody-based therapies that block PD-1 for use against lung cancer. Gene editing is expected to inhibit PD-1 with greater certainty, and by multiplying the cells, the scientists can increase the chance of triggering an immune response against tumours.

It is well known that CRISPR–Cas9 can result in edits at the wrong place in the genome, with potentially harmful effects. Chengdu MedGenCell, a biotechnology company in China and a collaborator on the trial, will validate the cells to ensure that the correct genes are knocked out before the cells are re-introduced into the patients, says oncologist Lei Deng of West China Hospital, a member of Lu's team.

Because the technique targets T cells — which are involved in various types of immune response — in a non-specific way, Chan worries that the approach might induce an autoimmune response in which T cells circulating in the blood might start to attack the gut, adrenaline glands or other normal tissue.

He suggests, instead, that the team take T cells from the site of the tumour because they would already be specialized for attacking cancer. But Deng says that the lung-cancer tumours targeted by their trial are not easily accessible. He also says that the team is reassured by the FDA-approved antibody therapies, which did not show a high rate of autoimmune response.

SAFETY FIRST

The phase I trial is designed foremost to test whether the approach is safe. It will examine the effects of three different dosage regimens on ten people, and, Deng says, the team plans to proceed slowly by increasing the dosage gradually and starting with just one patient, who will be monitored closely for side effects. But the researchers will also closely watch markers in the blood that would indicate that the treatment is working.

China has had a reputation for moving fast — sometimes too fast — with CRISPR, says Tetsuya Ishii, a bioethicist at Hokkaido University in Sapporo, Japan.

Lu says that his team was able to progress so rapidly because they are experienced with clinical trials of cancer treatments. The review process, which took half a year, included close communication with the hospital's internal review board (IRB). "There was a lot of back and forth," he says. The NIH's approval of the other CRISPR trial "strengthened our and our IRB's confidence in this study", he adds.

June is not surprised that a Chinese group has jumped out in front, he says, because "China places a high priority on biomedical research". Ishii notes that the clinical trial would be the latest in a series of firsts for China in the field of CRISPR — including the first CRISPR-edited human embryos and monkeys. "When it comes to gene editing, China goes first," he says.

Lu remains cautious. "I hope we are the first," he says. "And more importantly, I hope we can get positive data from the trial." ■

EUROPEAN UNION

Major funder tracks impact

European Research Council embarks on an unusual evaluation that could inspire others.

BY ALISON ABBOTT

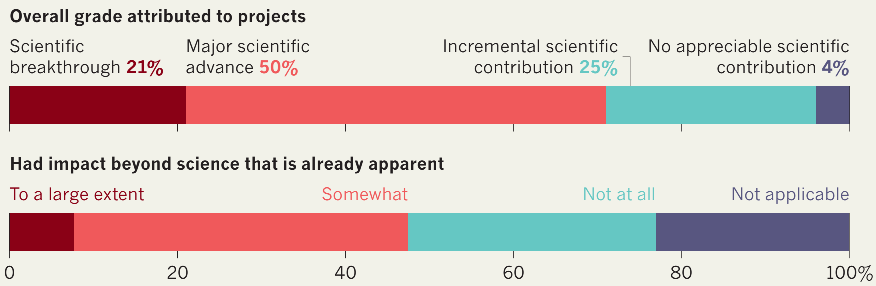
Last month, neuroscientist Ileana Hanganu-Opatz began a risky project with a risqué name: Psychocell. With a grant of €2 million (US$2.2 million), she is studying whether a single type of neuron causes a miswiring in the developing brain that has been linked to psychiatric disease. But it may turn out that no 'psychocell' exists, or that her mouse models are unsuitable.

Supporting such blue-skies research is the mission of her funder, the prestigious European Research Council (ERC), which launched in 2007 to raise the quality of European science. "No one but the ERC would have funded such a high-risk project," says Hanganu-Opatz, from the University of Hamburg, Germany.

Now, the council, which sits within the European Union's Framework funding programmes and has a €1.7-billion budget this year, has embarked on an unusual exercise: to retrospectively evaluate the success of the projects it funds. By contrast, most funding agencies assume that the evaluation to select which projects they fund is sufficient.

TO SCIENCE AND BEYOND

Scientific experts evaluated 199 completed projects funded by the European Research Council — including their contribution to science and to wider society.



SOURCE: ERC

"Virtually no basic research funding agency tries retrospectively to analyse its own performance and impacts," says Erik Arnold, chair of Technopolis, a European research and innovation consultancy headquartered in Brighton, UK. "It would be nice if the ERC effort would inspire others to do so."

On 26 July, at the European Science Open Forum in Manchester, UK, ERC president Jean-Pierre Bourguignon announced the results of a pilot investigation of 199 completed projects, almost three-quarters of which were deemed to have resulted in a scientific breakthrough or major advance (see 'To science and beyond').

"We push both scientists and grant-application reviewers to take a certain risk, so it is important to know that they are actually taking risks — and that we are selecting the right projects," says Bourguignon.

The ERC now plans to evaluate a selection of completed projects each year and to keep refining its methodology. Bourguignon hopes that this will help the council during discussions ▶

with politicians. “We want evidence-based arguments to show that bottom-up, curiosity-driven research is valuable to society,” he says. The council will have to lobby to keep its generous funding in the next Framework programme, due to begin in 2021.

The pilot evaluation rated projects that were among the first to be funded by the council, mostly in 2007 and 2008. It assigned eight projects each to 25 three-person expert groups.

The ERC gave the experts a bibliometric analysis of the publications from each project, but asked them to use their professional judgement to form an overall view of each one.

They found that 43 had led to a scientific breakthrough, 99 had generated a major advance — and only 7 had had no appreciable scientific output. That indicates an appropriate level of risk and ambition, says Bourguignon.

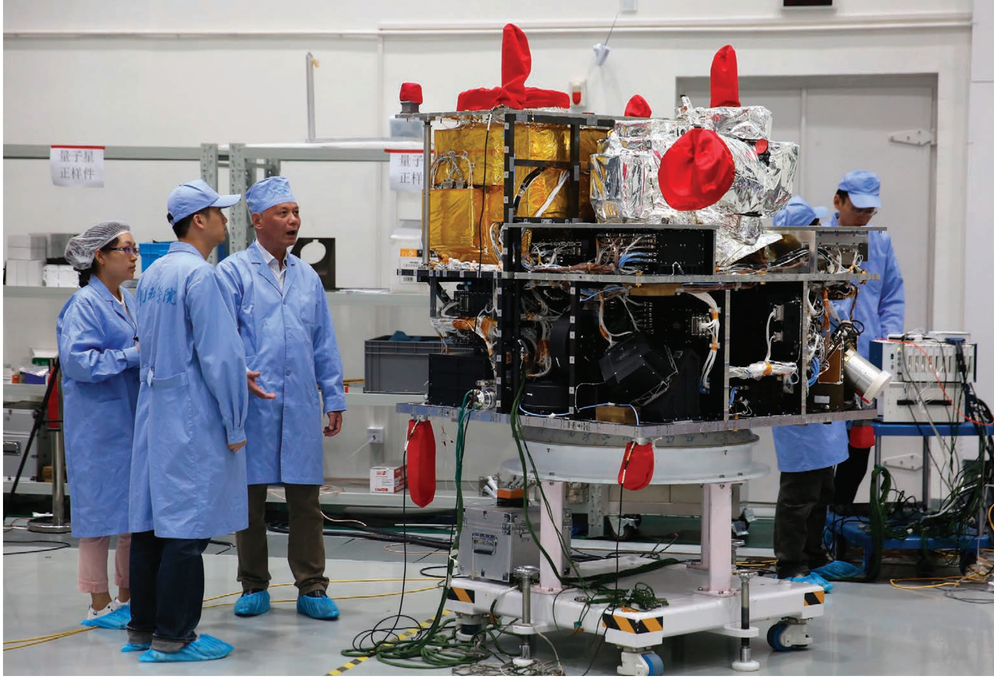
The evaluators also judged that almost 10% of projects had already had a large impact on the economy, policymaking or other aspects of society, and that around one-quarter were likely to do so in the future.

“It’s a delight to see a qualitative approach,” says science-policy specialist Ben Martin at the University of Sussex in Brighton. “Bibliometrics are misleading in isolation — but too often used this way.” Bourguignon says that many of the evaluators, who remain anonymous, struggled with the unfamiliar task of subjectively declaring research a “scientific advance”.

The study has limitations. Two experts in each group had served on ERC grant-awarding panels. None of the projects that they judged was included in the analysis, but the process could seem unobjective, says Arnold. Martin says the terms used to categorize the projects may be interpreted differently across disciplines. “As a social scientist, I can tell you that we don’t describe our work in terms of ‘breakthroughs’.”

Bourguignon agrees that the small study was not optimally designed. But the ERC has since solicited independent comments on the methodology, and an ongoing evaluation of a further 250 projects has been fine-tuned to let evaluators across disciplines report consistently.

In the pilot review, evaluators also stressed the ERC’s impact on an individual’s career, something that Hanganu-Opatz experienced at first hand. Her university gave her tenure on 18 July, and three other universities made her offers. “The visibility you get when you win an ERC grant is embarrassing,” she says. ■ SEE EDITORIAL P.465



China’s 600-kilogram quantum satellite contains a crystal that produces entangled photons.

CAI YANG/XINHUA VIA ZUMA WIRE

COMMUNICATIONS

One giant step for quantum internet

Chinese satellite is first in a wave of planned craft that could form network secured by quantum cryptography.

BY ELIZABETH GIBNEY

China is poised to launch the world’s first satellite designed to do quantum experiments. A fleet of quantum-enabled craft is likely to follow.

First up could be more Chinese satellites, which will together create a super-secure communications network, potentially linking people anywhere in the world. But groups from Canada, Japan, Italy and Singapore also have plans for quantum space experiments.

“Definitely, I think there will be a race,” says Chaoyang Lu, a physicist at the University of Science and Technology of China in Hefei, who works with the team behind the Chinese satellite. The 600-kilogram craft, the latest in a string of Chinese space-science satellites, will launch from Jiuquan Satellite Launch Center in August. The Chinese Academy of Sciences and the Austrian Academy of Sciences are collaborators on the US$100-million mission.

Quantum communications are secure because any tinkering with them is detectable. Two parties can communicate secretly — by sharing a encryption key encoded in the polarization of a string of photons, say — safe in the knowledge that any eavesdropping would leave its mark.

So far, scientists have managed to demonstrate quantum communication up to about 300 kilometres. Photons travelling through optical fibres and the air get scattered or absorbed, and amplifying a signal while preserving a photon’s fragile quantum state is

MORE ONLINE

TOP NEWS

Brexit watch: UK researchers scramble to save science go.nature.com/2arxw11

MORE NEWS

- Dutch agency launches first grants programme dedicated to replication go.nature.com/2a6qjwh
- Parasite infection calls zebrafish work into question go.nature.com/2a6qatn
- The Milky Way’s dark twin go.nature.com/2a6qh1b

NATURE PODCAST

Timing our breathing; working with indigenous peoples; and yeast genetics nature.com/nature/podcast

extremely difficult. The Chinese researchers hope that transmitting photons through space, where they travel more smoothly, will allow them to communicate over greater distances.

At the heart of their satellite is a crystal that produces pairs of entangled photons, whose properties remain entwined however far apart they are separated. The craft's first task will be to fire the partners in these pairs to ground stations in Beijing and Vienna, and use them to generate a secret key.

During the two-year mission, the team also plans to perform a statistical measurement known as a Bell test to prove that entanglement can exist between particles separated by a distance of 1,200 kilometres. Although quantum theory predicts that entanglement persists at any distance, a Bell test would prove it.

The team will also attempt to 'teleport' quantum states, using an entangled pair of photons alongside information transmitted by more conventional means to reconstruct the quantum state of a photon in a new location.

"If the first satellite goes well, China will definitely launch more," says Lu. About 20 satellites would be required to enable secure communications throughout the world, he adds.

The teams from outside China are taking a different tack. A collaboration between the National University of Singapore (NUS) and the University of Strathclyde, UK, is using cheap 5-kilogram satellites known as cubesats to do quantum experiments. Last year, the team launched a cubesat that created and measured pairs of 'correlated' photons in orbit; next year, it hopes to launch a device that produces fully entangled pairs.

Costing just $100,000 each, cubesats make space-based quantum communications accessible, says NUS physicist Alexander Ling, who is leading the project.

A Canadian team proposes to generate pairs of entangled photons on the ground, and then fire some of them to a microsatellite that weighs less than 30 kilograms. This would be cheaper than generating the photons in space, says Brendon Higgins, a physicist at the University of Waterloo, who is part of the Canadian Quantum Encryption and Science Satellite (QEYSSat) team. But delivering the photons to the moving satellite would be a challenge. The team plans to test the system using a photon receiver on an aeroplane first.

An even simpler approach to quantum space science, pioneered by a team at the University of Padua in Italy led by Paolo Villoresi, involves adding reflectors and other simple equipment to regular satellites. Last year, the team showed that photons bounced back to Earth off an existing satellite maintained their quantum states and were received with low enough error rates for quantum cryptography (G. Vallone *et al. Phys. Rev. Lett.* **115,** 040502; 2015).

Researchers have also proposed a quantum experiment aboard the International Space Station (ISS) that would simultaneously entangle the states of two separate properties of a photon to make teleportation more reliable and efficient.

As well as making communications much more secure, these satellite systems would mark a major step towards a 'quantum internet' made up of quantum computers around the world, or a quantum computing cloud, says Paul Kwiat, a physicist at the University of Illinois at Urbana–Champaign who is working with NASA on the ISS project.

Eventually, quantum teleportation in space could even allow researchers to combine photons from satellites to make a distributed telescope with an effective aperture the size of Earth — and enormous resolution. "You could not just see planets," says Kwiat, "but in principle read licence plates on Jupiter's moons." ■

CORRECTION

The News story 'Canada builds quake warning system' (*Nature* **534,** 446–447; 2016) incorrectly stated that the warning system being developed by Ocean Networks Canada would be the first in Canada.

The ice historians

TO TELL WHETHER GREENLAND'S GLACIAL CAP WILL MELT AWAY ANY TIME SOON, RESEARCHERS ARE PORING OVER OLD PHOTOGRAPHS AND DRAWINGS FOR CLUES TO ITS PAST BEHAVIOUR.

BY QUIRIN SCHIERMEIER

ARCTIC INSTITUTE

Danish geologist Lauge Koch (centre) and colleagues on the open-cockpit Heinkel plane that his team used to survey East Greenland in 1932.

Anders Bjørk could have stepped out from a painting of a nineteenth-century Arctic expedition. Tall, athletic and burning for outdoor action, he has been attracted to the wild fjords and glaciers of Greenland since his university days, when he took summer tourists on guided tours there. He still ventures to the island every summer to measure its waning ice sheet — and he doesn't always travel by air. In 2014, he retraced the route of early Danish explorers by sailing to Greenland in a three-masted wooden schooner built in the 1930s.

But on a pleasant spring day in Copenhagen, Bjørk is carrying out a much safer reconnaissance. In the belly of Denmark's national archive, he sifts through yellowed files filled with rows of daily weather and ocean-temperature measurements that Danish clergymen and village chiefs made decades ago along the Greenland coast.

"I get a kick out of sitting over old maps and documents," he says. "I somehow find that kind of stuff just as fascinating as doing field work." Bjørk has alternated expeditions to the archives and to distant glaciers for ten years, ever since he first came across tens of thousands of aerial photographs of Greenland from the 1930s, lying all but forgotten in a seventeenth-century fortress outside Copenhagen.

Bjørk and his colleagues at the Natural History Museum of Denmark in Copenhagen are ice historians. They are combing the old records to document how Greenland's ice sheet and glaciers have behaved since the nineteenth century — a crucial set of information for climate scientists trying to predict how it might change in the future.

With Arctic temperatures rising faster than anywhere else on Earth, Greenland is now losing about 200 billion tonnes of ice per year and raising ocean levels around the globe. Projections suggest that melting from the island might swell sea levels by 30 centimetres by the end of this century. If all Greenland's ice melted — a possibility over the next few centuries – it would push up sea level by more than 6 metres, enough to flood coastal megacities such as New York and Miami.

But the projections carry large uncertainties, in part because researchers lack basic information about Greenland's past. Satellite data only go back 40 years, which is why Bjørk and his colleagues are poring over 180,000 photographs and other data that record how glaciers have advanced and retreated during cold and warm spells in the recent past. Their first sets of findings suggest that Greenland ice has responded more strongly to past climate changes than was previously realized. Now, the researchers are trying to unravel what factors within the oceans, atmosphere and inside glaciers control their behaviour.

What sets Bjørk apart, say other scientists, is that he combines the heart of a seafarer with a strong sense of detail and creativity in research. The studies by him and his fellow ice historians are making key contributions to glacial science, says Beata Csatho, a glaciologist at the University of Buffalo in New York. "Collecting historical data sets is probably more important at this point than having yet another satellite do more of the same stuff," she says. "This is not incremental progress — it is a big advance in science."

EPIC ADVENTURES

Bjørk's fascination with Arctic exploration reaches back to his childhood. Like many Danes, he learned to sail at a young age and over the years he embarked on a series of increasingly epic voyages that blended science and adventure.

In Bjørk's most extreme expedition, he and a small crew of like-minded scientists sailed through the storm-tossed south Atlantic for two months in 2011. Svante Björck, a Quaternary geologist at the University of Lund in Sweden, had chartered a two-masted ship to sail from the Falkland Islands to Tristan da Cunha, the most remote inhabited island in the world. During the 13-day trip, Bjørk spent several nights awake on deck, tied to the mast during gales, helping the captain steer the ship through huge waves. "It was scary. I've never seen anything like this," he says. The return journey to Uruguay took 33 days against relentlessly unfavourable trade winds. Yet the rich yield in climate data and the unforgettable adventure was well worth the ordeal, he says.

The son of two school teachers, Bjørk grew up in Svendborg, a small harbour town in southern Denmark with a long seafaring tradition. Eager to explore the world, he studied geography and became enthralled with Greenland, which has been part of the kingdom of Denmark since the early 1800s. While serving as a student assistant to Quaternary geologist Kurt Kjær of the Natural History Museum of Denmark, Bjørk learned to mix traditional geology and expedition-era geography with modern high-tech data.

His interest in historic records was piqued ten years ago when he first heard tales of a cache somewhere in Copenhagen filled with aerial images of Greenland taken in the 1930s during mapping expeditions. He eventually discovered that the formerly classified images had been locked away in a ramshackle citadel outside the city. Bjørk sensed a unique research opportunity, but it took him years to sift through rows and rows of boxes filled with thousands of light-sensitive negatives and to work out which might be useful.

As he extended his search, he came across many more images taken

during Danish–US Greenland aerial surveys from the 1940s to the 1970s. Then came pictures from Danish mapping expeditions carried out from 1978 to 1987. Bjørk has now amassed a collection of 180,000 aerial photographs (see 'Flight lines'). But he has paid particular attention to the first images he acquired, from the 1930s, because they chronicle a warm period that shares some similarities with the present. The mapping expeditions, led by the explorer Knud Rasmussen, were partly motivated by a dispute with Norway about which country had sovereignty over Greenland — a legal battle that the international court in The Hague settled in 1933 in favour of Denmark. But the crews were aware of the scientific potential of their work. "The instantaneous record contained in the air and terrestrial photographs of the present state of the glaciers of the entire coast is in itself of great value for future researchers," the Danish captain Carl Gabel-Jørgensen wrote in a 1935 report.

"WITHOUT GPS OR OTHER PRECISION TOOLS THEY STILL MANAGED TO PRODUCE GLACIER DATA OF AMAZING QUALITY."

"These men were professionals," says Eric Rignot, a glaciologist at the University of California, Irvine, whose group Bjørk will join in the autumn as a postdoc. "Without GPS or other precision tools they still managed to produce glacier data of amazing quality."

The image collection from the 1930s includes 10,000 overlapping aerial photographs captured with a rotating camera flown in an open hydroplane at 4,000 metres altitude. Bjørk and his colleagues have used these images — alongside more recent aerial images and satellite observations — to produce an 80-year record of how 132 glaciers in south-east Greenland have waxed and waned[1]. They found that the glaciers retreated strongly during both warm periods but many have lost more ice during the record temperatures in recent decades.

Not all glaciers have responded to warming in the same way. It has hit low-elevation glaciers harder than ones higher up, and glaciers that terminate in the sea seem to be more vulnerable to the present rise in temperatures than those that end on land. Researchers are particularly concerned about how quickly these marine glaciers might shrink as they are attacked by the rapidly warming ocean. That may be a key factor in determining how quickly Greenland loses ice.

The signs are not good. Hundreds of outlet glaciers that drain the ice sheet are losing substantially more mass and are thinning more than they are gaining through snow accumulation. Jakobshavn, one of Greenland's largest glaciers, is moving faster than all the others and is the poster child for the fragile ice sheet, which some fear might approach a fatal tipping point as air temperatures continue to rise and the glaciers thin even more[2].

The findings from the historical studies released so far are a reminder that the current glacier retreat has lasted far longer and is more pronounced than the typical fluctuations over the past century. "It's that kind of long-term perspective that helps us understand that the rapid changes we are seeing now are definitely more than just noise," says Csatho, who has looked at satellite and aerial data to reconstruct how the surface elevation of the Greenland ice sheet changed at nearly 100,000 locations from 1993 to 2012 (ref. 3).

RECOVERED TREASURE

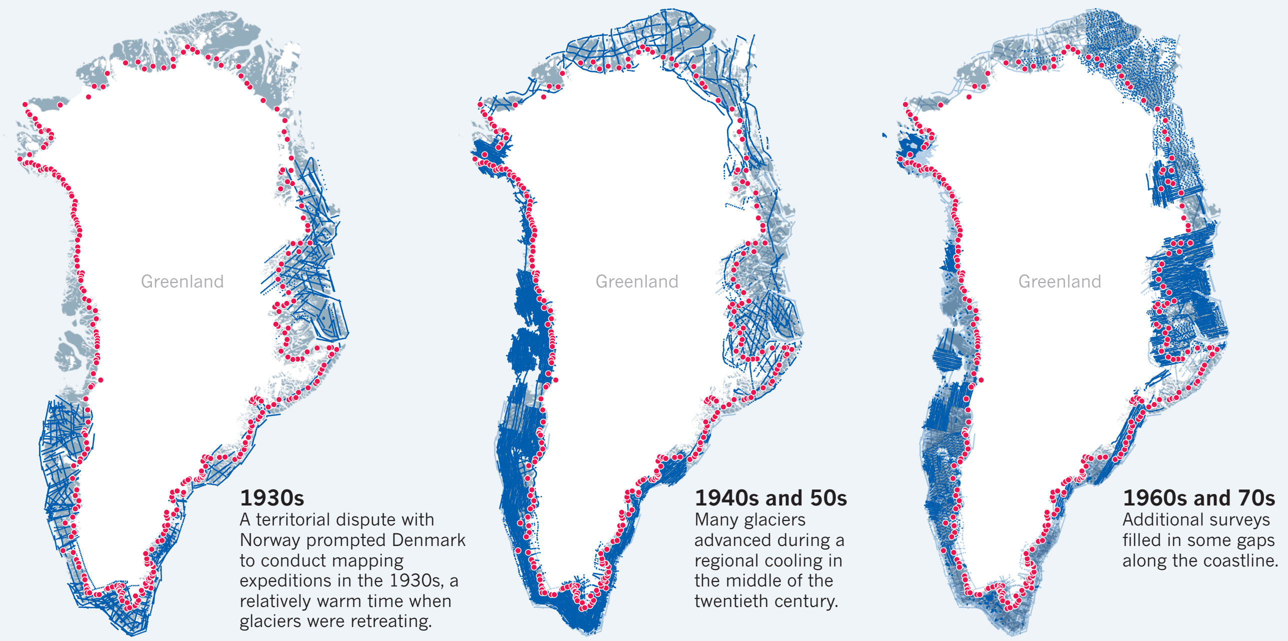
Bjørk's work at the Arctic Institute in Copenhagen is more sedate than his trips to Greenland, but no less rewarding. The centre's archives hold treasured documents and memorabilia of Denmark's Greenland expeditions, which are part of the Nordic kingdom's cultural DNA. Bjørk gleams with the delight of an art lover as he unrolls maps drawn by Danish explorer Hinrich Rink in 1851 of the mighty Jakobshavn. Carefully opening stitched albums, Bjørk passes by snapshots of Inuit villagers with sealskin boats, and heads for the pages that most interest him — the sketches and images depicting the shapes of glaciers and their thickness.

"Rink and later explorers documented very carefully where they were, and we can trust that what they painted is exactly what they saw," he says. "Their zeal and accuracy is a gift for us."

Bjørk has done much of his work with fellow geographer Kristian Kjedlsen, with whom he shares an office in the Natural History Museum plastered with Greenland maps and Arctic photography. In a second study, they and their colleagues carefully picked out geographical features on old photographs, such as moraines of rock or vegetation lines

FLIGHT LINES

Anders Bjørk and his colleagues are using historical aerial surveys to assess Greenland's glaciers, including 309 major ones (•) that drain the ice sheet. Blue lines represent flight tracks.

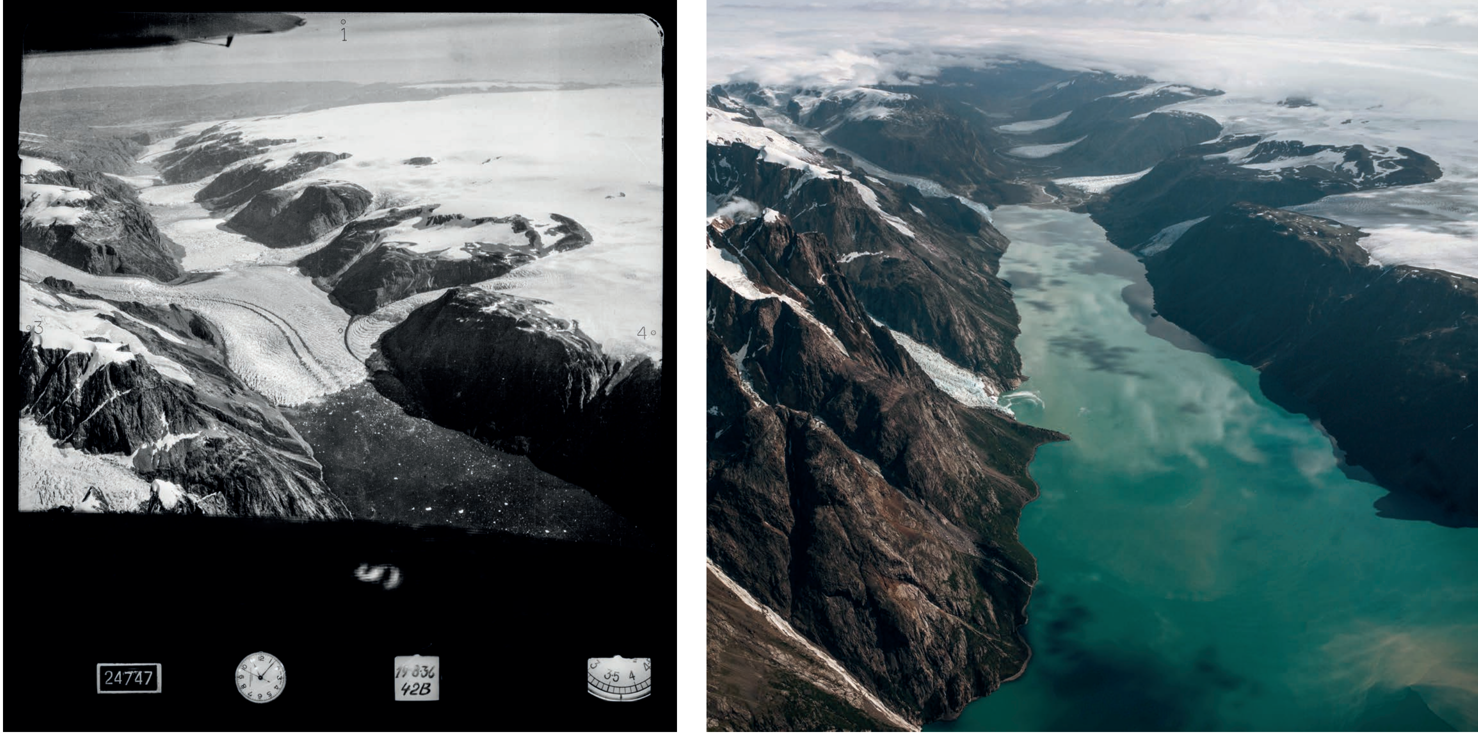


1930s A territorial dispute with Norway prompted Denmark to conduct mapping expeditions in the 1930s, a relatively warm time when glaciers were retreating.

1940s and 50s Many glaciers advanced during a regional cooling in the middle of the twentieth century.

1960s and 70s Additional surveys filled in some gaps along the coastline.

L: DANISH GEODATA AGENCY; HANS HENRIK THOLSTRUP/
R: NAT. HIST. MUS. DENMARK



Photographs of Ujaraannaq Valley in southwest Greenland in summer in 1936 (left) and in 2013 show that several glaciers have disappeared.

that mark the greatest extent of the glaciers during the Little Ice Age at the end of the nineteenth century.

They used these data, along with more recent aerial surveys, to determine how the height of the glaciers has shifted, which provides a way to track changes in the ice mass. Their results indicate that the mean annual rate of ice loss between 2003 and 2010 was more than double the average for the twentieth century[4]. Moreover, striking spatial differences seem to confirm that the ice sheet's response to changing climates is governed by the topography of the underlying bedrock and by the geometry of the fjords through which outlet glaciers are flowing towards the sea.

UNRULY ICE

Greenland's glaciers have been fickle over the past century, advancing in some places and retreating in others. By discovering when those changes happened and what kind of conditions prevailed at the time, Bjørk and his co-workers hope to shed light on the ice sheet's complicated mechanics — a missing piece in attempts to model its waxing and waning accurately.

But nailing the ice sheet's behaviour requires looking at as many individual glaciers as possible. This is just what Bjørk and his group are doing now. Using all available historic records from Greenland, including pictures of glaciers in the least-explored high north, alongside modern satellite imagery, he aims to reconstruct the history of the island's 309 biggest glaciers in unprecedented detail.

The work is in full swing. On most days, two of his students toil away in the basement of the museum in front of computers, clicking on boulders, cliffs and other recognizable features that can be spotted in photographs taken decades apart. These are control points, which allow the researchers to match pictures captured at different altitudes and angles. Most aerial pictures were not taken looking straight down and hence lack a single scale. All of these oblique pictures need to be converted, one by one, to vertical views so that they can be transferred onto a common coordinate system. Without doing this geo-rectification, the researchers could not accurately measure the glaciers' advances and retreats.

Photos are not the only source of information. In the Arctic Institute on a sunny afternoon, Bjørk looks at sketches that the German geologist Alfred Wegener produced during his last expedition, shortly before he died in November 1930 near an ice camp in central Greenland. Other albums hold illustrations of glaciers that scientists made during field trips in a cold spell in the late nineteenth century.

So far, he has discovered about 600 sketches and paintings that might help to tell the story of Greenland's glaciers before aerial photographs first became available. For some glaciers, official aerial pictures don't exist. To fill the gaps, he is also consulting privately held images that geologists have taken over the years. "New information is coming up from all sides," he says.

In 2014, the daughters of two pilots involved in the 1930s surveys offered Bjørk their large collections of photos and even an 8-millimetre film their fathers had taken during the campaigns.

The emerging story is that ice disappeared very fast in the early twentieth century in the warming that followed the end of the Little Ice Age. Then the subsequent cool spell brought widespread glacial advances. The profound switch between shrinking and surging suggests that the glaciers are more sensitive to warmings and coolings than researchers had previously thought, says Bjørk. But why some glaciers advanced forcefully at given periods and temperatures whereas others did not is still a puzzle.

NASA has launched the Oceans Melting Greenland (OMG) project, led by Rignot, to provide glaciologists and ice-sheet modellers with unprecedented base maps of fjord bathymetry and other information needed to determine how glaciers interact with the sea. That is where one of Bjørk's latest discoveries could prove useful. The historic records of sea-surface temperature that he unearthed can be combined with the individual histories of different glaciers to see how those that end in the ocean responded to changing marine conditions. The past behaviour of the ice, Rignot says, "matters a great lot when it comes to projecting its fate".

After all, the ultimate goal of this historical research is to look forward. Along with palaeoclimatic data from hundreds and thousands of years ago, the findings of Bjørk and his colleagues from the recent past promise to increase confidence in the projections of ice-sheet models, says Richard Alley, a glaciologist at Pennsylvania State University in University Park. "We need history as well as modern observations to build and test predictive models."

For Bjørk, the historical research goes well beyond science. It also connects him with the pioneering scientists and explorers he grew up admiring. He is grateful that their legacy is finally being dug out from the crypt. "It's part of Nordic history," he says, "and a real gift to modern science." ■

Quirin Schiermeier *writes for* Nature *from Munich, Germany.*

1. Bjørk, A. A. *et al. Nature Geosci.* **5,** 427–432 (2012).
2. Gregory, J. M., Huybrechts, P. & Raper, S. C. B. *Nature* **428,** 616 (2004).
3. Csatho, B. M. *et al. Proc. Natl Acad. Sci. USA* **111,** 18478–18483 (2014).
4. Kjeldsen, K. K. *et al. Nature* **528,** 396–400 (2015).

JUSTIN JIN FOR *NATURE*

The BEER GEEKS

BY EWEN CALLAWAY

A lab in Belgium is using genetics to make the perfect beer yeast.

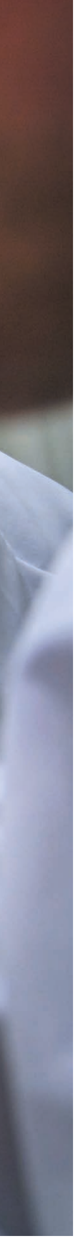
Kevin Verstrepen's lab meetings can be pretty boozy affairs. Twice a week, several members of his group at Belgium's University of Leuven and the Flanders Institute for Biotechnology gather around a table loaded with black, tulip-shaped beer glasses, together with spit buckets and crackers.

Verstrepen holds a glass and takes a whiff of its contents. "For me, this was an ethyl acetate bomb," he pronounces, referring to a chemical found in pear-flavoured sweets that, at high concentrations, reeks of nail polish.

Brigida Gallone, a graduate student in the lab, detects a second aroma: "ethyl acetate and 4-VG", she says. That's 4-vinyl-guaiacol, which smells of smoke, cloves and — according to a tasting sheet in front of her — a dentist's surgery. "I like 4-VG and this was too much for me." Another student, Stijn Mertens, catches a smell of wet cardboard, which is common to stale beers. "I got some *trans*-2-nonenal," he says. With that, the group finishes its analysis of this brew and moves onto the ninth and final glass. It is not even 11 a.m..

"There's only so many you can do before you lose focus," says postdoc Miguel Roncoroni, who has been hosting these tastings for more than 4 months. They are part of a project to characterize some 200 commercially produced Belgian beers. Their assessments, alongside precise measurements of the dozens of chemicals that produce the flavours and aromas, could help consumers to identify new beers to try, by comparing the lab's profiles to ones they like.

But Verstrepen has loftier ambitions than helping beer lovers to select



Kevin Verstrepen (centre) leads a beer-tasting session in his lab at the University of Leuven in Belgium.

their next bottle. He wants to build the perfect yeast. His lab is deploying what it is learning about the chemical and genetic basis of beer flavour to breed yeast strains that generate unique flavours and other qualities coveted by brewers and drinkers.

The beer geeks in his lab straddle the worlds of curiosity-driven science and industrial brewing. They study evolution, biochemistry, and even neuroscience through yeast. But they also have contracts with beer makers worldwide, from multinational conglomerates to small trend-setting craft breweries. In an upcoming *Cell* paper, the lab will report the genomes of some 150 yeast strains used to make beer, sake and other fermented products, a project done in collaboration with a leading supplier of yeast to brewers and a synthetic-biology firm.

For a US$500-billion industry whose products depend on complex interactions between chemistry and microbiology, sophisticated yeast strains are hot commodities. “You always want to know what’s new in Kevin’s lab,” says Peter Bouckaert, the brewmaster at New Belgium, a leading craft brewery in Fort Collins, Colorado. “People are watching what he does.”

Beer gets its flavour from just a few ingredients (see ‘Better brewing through biology’). Grains — mainly malted barley — provide sugar and body, but can also imbue flavour, such as the chocolate notes common to dark stouts. Hop flowers bring bitterness as well as the tropical fruit notes in some craft beers. Dissolved minerals in the water influence the flavours that come through from grain and hops. And brewer’s yeast, *Saccharomyces cerevisiae*, provides alcohol, bubbles and hundreds of aroma compounds. Fermentation produces everything from isoamyl acetate, which makes German Hefeweizens taste of banana, to the clove notes of 4-VG.

Yeast science was spearheaded by beer makers. Denmark’s Carlsberg brewery established one of the world’s first yeast-biology labs in 1875, and it was there that Emil Christian Hansen isolated the first pure culture of a brewing yeast in 1883. In the 1930s and 40s, another Carlsberg scientist, Øjvind Winge, discovered that yeast reproduces both sexually and asexually and used this insight to breed new strains with useful brewing traits.

Winge’s work moved yeast from the brewery to biology labs, and many scientists now use brewer’s yeast as a model to probe the inner workings of complex cells. But despite a long and fruitful marriage between yeast and bioscience, Verstrepen argues that many brewers are still stuck in the nineteenth century when it comes to the yeast they use. “Brewers, especially traditional brewers, are often not using the optimal yeast.” Most use just one strain — isolated from their brewery or borrowed from another decades ago — in all their beers.

Verstrepen wants to change that. He started out working in a South African wine yeast lab, then joined a University of Leuven beer lab for his PhD in 1999. But he was disappointed to find that most of the research involved troubleshooting brewers’ problems. “Nobody was doing any biology, really,” he says. Disillusioned, he moved to the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, to pursue a postdoc with Gerald Fink, who pioneered genetic engineering in yeast in the 1970s.

But although the scientists there liked yeast, nobody was interested in beer — at least not by day — Verstrepen says. So, his work centred on the proteins that pathogenic yeast deploy to stick to human tissue. Verstrepen discovered that the level of stickiness is dictated by the number of repeating DNA sequences in a certain gene (K. J. Verstrepen *et al. Nature Genet.* **37,** 986–990; 2005). “It’s like having longer Velcro loops that stick more easily,” he explains. The proteins are also responsible for flocculation, the process by which yeast cells clump together in a beer and fall out of solution. Flocculation varies between brewing strains and influences the flavour, clarity and alcohol content of a beer.

In 2005, Verstrepen moved across town to open a lab at Harvard University, focusing on the roles of various repetitive DNA sequences in generating diversity. He taught Harvard undergraduates biology in a course that incorporated brewing — “It was a pretty tough course,” he says — but beer didn’t feature in his research until he returned to Leuven in 2009.

Verstrepen hoped to combine the kind of research he was doing with his interests in beer and wine. But it was a phone call from a Swiss chocolate company that jump-started his work with industry. Zurich-based Barry Callebaut, one of the world’s largest cocoa makers, needed help transforming bitter cacao beans into cocoa powder (which is traditionally done by yeasts present in the environment). “My response was ‘Is chocolate fermented?’ That wasn’t the smartest thing to say,” says Verstrepen.

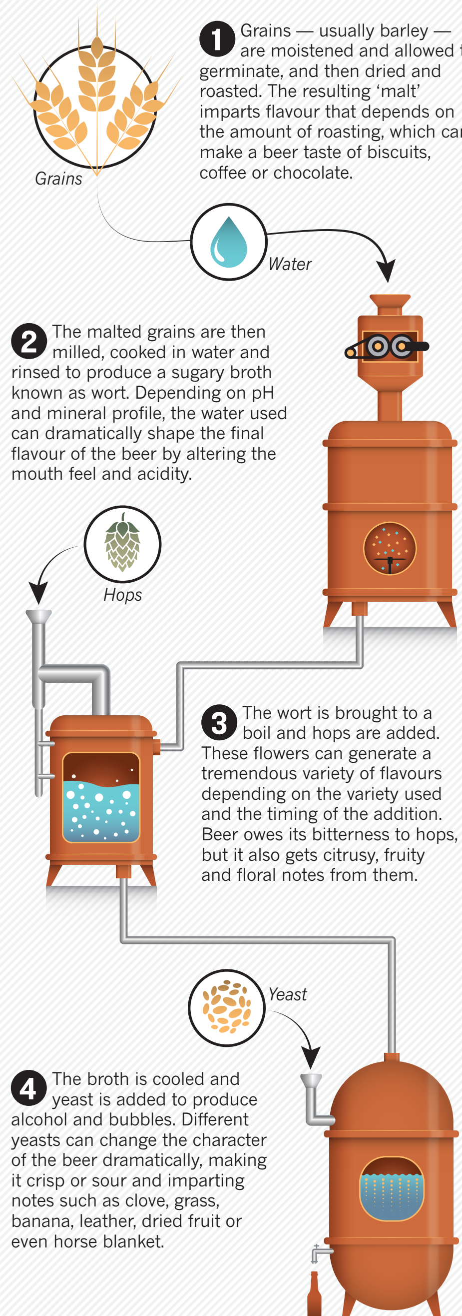
Nevertheless, the company became the first client of Verstrepen’s lab consultancy, which now earns around €500,000 ($555,000) per year. Half of the 25 or so scientists in his lab do applied research on beer, biofuels and other fermented products; the rest pursue epigenetics, molecular evolution and other basic research.

LAGER LAB

At first glance, the lab looks like any other. Centrifuges, petri dishes and pipettes cover benches. There is also an incubator full of small glass bottles. It would look at home in any microbiology lab, were the bottles not filled with a strong broth of malted barley, sugar and hops. The cold room offers more clues. “There’s not too much beer in here now,” Verstrepen says, pointing to a few full crates. And a stockroom contains hundreds of the black glasses used for blind tastings; they are opaque so tasters cannot see what they are sipping.

BETTER BREWING THROUGH BIOLOGY

Four ingredients generally influence the flavour and aroma of beer.



1 Grains — usually barley — are moistened and allowed to germinate, and then dried and roasted. The resulting ‘malt’ imparts flavour that depends on the amount of roasting, which can make a beer taste of biscuits, coffee or chocolate.

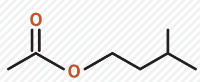
2 The malted grains are then milled, cooked in water and rinsed to produce a sugary broth known as wort. Depending on pH and mineral profile, the water used can dramatically shape the final flavour of the beer by altering the mouth feel and acidity.

3 The wort is brought to a boil and hops are added. These flowers can generate a tremendous variety of flavours depending on the variety used and the timing of the addition. Beer owes its bitterness to hops, but it also gets citrusy, fruity and floral notes from them.

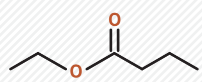
4 The broth is cooled and yeast is added to produce alcohol and bubbles. Different yeasts can change the character of the beer dramatically, making it crisp or sour and imparting notes such as clove, grass, banana, leather, dried fruit or even horse blanket.

TAPPING YEAST

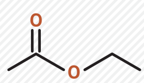
By-products of fermentation produce a wide variety of flavours and aromas, and their levels can be tuned by breeding new yeast strains.



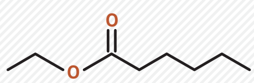
Isoamyl acetate: banana – probably the most prominent ester in beer



Ethyl butyrate: tropical fruit, bubble gum

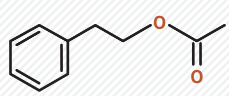


Ethyl acetate: pear, nail varnish



Ethyl hexanoate: red apple

Acetaldehyde: green apple



Phenylethyl acetate: flowers

His lab's freezers house about 30,000 types of yeast, including 1,000 strains used by brewers, bakers and others worldwide, and another 1,000 wild isolates from fruit, flowers, insects and even people. Many of these have been characterized for the genes known to influence taste and other traits that brewers care about. The lab is collaborating with yeast supplier White Labs in San Diego, California, and with Craig Venter's Synthetic Genomics in nearby La Jolla to build a family tree of industrial yeast.

The balance in the freezers is made up by the lab's creations — completely new strains that have unique combinations of traits. The team makes them by mating different strains and then screening the offspring for the aromas they produce and, more recently, for genes that underlie these traits. This latter approach, known as genetic-marker-assisted breeding, is common in agriculture, and Verstrepen thinks that it will transform brewing.

A Canadian brewery commissioned one of these custom strains when it wanted a full-flavoured Belgian 'tripel' that has less alcohol than is typical for the style. Another brewery has asked for yeast that makes chocolatey aromas, a request that has stumped the lab so far.

MASS MATING

Using a robot that can accomplish hundreds of matings in a day, the lab generates many more strains than it can analyse or use in its taste tests. To cope with the glut, the researchers are developing microfluidic chips that can brew with 2,000 different yeasts at a time in 20-picolitre batches, each of which contains a single yeast cell. They can automatically test the resulting 'picobrews' for alcohol content and hope eventually to be able to measure the aroma compounds produced, too. At the other end of the spectrum, the lab will soon take delivery of a kit to make 500-litre batches of beer to better appreciate the challenges of industrial-scale brewing.

Verstrepen's archive makes the lab a one-stop shop for brewers looking for unique flavours. For example, when Bouckaert wanted "funky" aromas, such as the smell of barnyards or horse blankets, but without the usual accompanying fruity notes, he tried four of the lab's strains that showed some promise. None was going to work for a New Belgium beer — at least not yet. "Kevin's research is a little bit out there and on the edges of brewer's applications," Bouckaert says. "But that doesn't mean it will not translate to something that could be huge in the future."

Natural variation in brewer's yeast offers much leeway to dial up and down flavours and other traits, but the approach can only go so far, Verstrepen says. Genetic-modification tools could improve on that. "With breeding we can increase flavour 10-fold, with genetic modification we can increase it 100-fold or 1,000-fold, and we've done this," Verstrepen says. "The beers you make are more like banana milkshakes. Is this something we really want to do?" Brewers are excited about the work, but the stigma that surrounds genetically modified (GM) foods means that the lab always uses more conventional techniques such as breeding and directed evolution to make strains destined for industry.

And gene-editing techniques such as CRISPR could be used to introduce naturally occurring variants linked to flavour production into yeast strains that perform well but don't impart much flavour. This would accomplish the same goals as conventional breeding much more quickly. Food regulators are already grappling with the question of whether plants and livestock made this way count as GM foods, so it's not inconceivable that CRISPR beers could force the issue.

"THE BEERS YOU MAKE ARE MORE LIKE BANANA MILKSHAKES. IS THIS SOMETHING WE REALLY WANT TO DO?"

A few craft breweries have asked the lab for GM yeast (they were turned down), but most of the brewing industry has little appetite for it, says Bouckaert. "American craft is pushing the boundaries, but on genetic modification, it's a no-no," he says. Multinational brewers are even more skittish about being linked to GM beer.

Leuven-based AB InBev, the world's biggest brewing conglomerate and the maker of Budweiser and Stella Artois, confirmed that it works with Verstrepen. Philippe Malcorps, a senior scientist in the company's yeast and fermentation division, says that AB InBev is interested in tapping the diversity of wild yeasts to get flavours and other traits that would be impossible to find in classic brewing strains.

Much of Verstrepen's work with such companies is covered by strict non-disclosure agreements. "Some of our best applied research, we cannot publish, we cannot talk about," he says, "That is frustrating as a scientist, because you want to share."

But smaller breweries tend to be more liberal. In fact, Orval, one of the six Belgian breweries run by Trappist monks, declined to sign a non-disclosure agreement or even a contract when it asked Verstrepen to sequence its house strain. "If you cannot trust monks any more, what are you going to do?" says Verstrepen (who keeps his personal refrigerator stocked with Orval, one of his favourite beers). Bouckaert says that he was offered exclusive access to the yeast strains his company was trying, but he balked. "I said, 'I'm a craft brewer. I don't want exclusivity. I want you to send them to more people!'"

The applied work has led to scientific insights. Verstrepen's team is still grappling with a project that originated from trouble-shooting a problem it heard about from brewers. Often, brewers reuse yeast to cut costs. But periodically, they start new cultures, growing them first on glucose (yeast's preferred food), then adding them to unfermented beer, which contains mainly maltose. Yeast sometimes responds sluggishly to the new food source, and it can take days before the beer is fermenting at full clip (and that leaves it prey to contaminants). "Yeast remember not only what they were eating, but also what their grand-grand-grandmothers were eating, up to five or six generations," Verstrepen says. Working out the mechanism, which seems to involve the inheritance of epigenetic modifications to DNA or proteins, is now a focus of the lab.

Often, the lab has found, good beer and good science go hand in hand. For his PhD project, Mertens was tasked with creating new strains of the yeast used to make lagers. Lager is brewed at cooler temperatures than other beers, using a yeast that emerged several centuries ago when *S. cerevisiae* hybridized with a related cold-tolerant species called *Saccharomyces eubayanus*.

Although responsible for the vast majority of beer sales, lagers tend to have a limited flavour range. To make yeasts that might expand this palette, Mertens coaxed various strains of the two species into mating. Some beers made with these yeasts tasted of onion, and many fermented poorly, but one combined the bracing crispness of a good pilsner (a pale lager) with hints of fruity aromas that are untypical of the style (S. Mertens *et al. Appl. Environ. Microbiol.* **81,** 8202–8214; 2015). The beer was so popular among lab members that bottles vanished from the cold room. Several breweries are now testing that strain, known as H29.

Mertens would love to see a commercial beer made from his creation, but he also hopes to sequence the genomes of the other yeasts he has created to work out how species hybridize — and maybe even unpick the conditions that spawned the first lager yeast. "We have new yeast and the brewers love it," he says. "But we're looking at the fundamentals of how these hybrids work. That goes beyond beer science."

By the end of the morning's beer tasting, the spit buckets are filling up. Verstrepen has a meeting with a DNA-sequencing company, and Mertens and the other students have their research projects to get to. The lab might be a magnet for beer geeks, but it is no keg party.

"Yes, you're working with a very fun product, but at the end of the day, it's genetics work," Mertens says. "We're not drinking for fun." At least, not until after work. ■

Ewen Callaway *writes for* Nature *from London.*

COMMENT

ILLUSTRATION BY DAVID PARKINS

Lessons from Brexit

Britain's vote to leave the European Union is jeopardizing scientists' funds, collaborations, staff and students; it has left the nation reeling and Europe vulnerable. These schismatic times have researchers worldwide soul-searching over how best to contribute. Five experts offer their reflections.

JOHAN SCHOT

Rethink social progress

Director of Science Policy Research Unit, University of Sussex

The popular rebuke to reason that was Britain's vote to leave the European Union is a wake-up call. Our world requires an urgent rethinking of social progress. The sciences, social sciences and humanities should collaborate and open up their research agendas for public engagement and interdisciplinary dialogue to work towards a diversity of possible solutions to address the troubles of our time.

In the debate about the consequences of Brexit, the EU is often conflated with Europe. But the start of international governance with a strong European outlook came long before — during the nineteenth century — with industrialization and globalization. The EU is hence part of a greater 'Europeanization' process involving a web of multiple organizations and dependencies. Leaving the EU does not mean leaving Europe.

The deliberations about Europe's future should therefore look more widely to the crucial question of how to organize and manage the international flows of goods, people, information, pandemics and pollution — inevitable in our globalized world.

Many have, rightly, criticized the EU for its technocratic character. Yet this is present in all international organizations — from the United Nations and the Council of Europe to the European Space Agency and ▶

CERN, European's particle-physics lab. These bodies have been built on the belief in the rule of experts — people like us — our rational decision-making and, if we're honest, our preference for avoiding overt political and public deliberation.

It is time to reappraise these axioms of both European and global governance. Current designs reflect the old dream of building a European, or even global, state that would eventually surpass nation-states. Public support for this is clearly faltering: many issues need localized solutions. New constructs for local, national and international governance are needed that combine technocracy and democracy to reduce the democratic deficit of which Brexit is a symptom.

This is crucial if we are to address challenging interconnected issues such as rising inequalities, unemployment, growing migration, climate change, human rights, arms control, cyber-insecurity and terrorism. Brexit must not simply prompt organizations in Europe to engage in yet another round of navel-gazing discussion focused on the future of their own interests, particularly funding and trade.

"We should encourage experimentation with a variety of new democratic models."

Instead, we should encourage experimentation with a variety of new democratic models at all levels to create more participative forms of policy-shaping. We should focus less on stimulating research, development and innovation across the board, as if it were all positive. Instead we must open up the debate about how knowledge production contributes to social progress to address the societal and environmental challenges of the twenty-first century.

LEVENTE LITTVAY

Compare populist movements

Associate professor of political science, Central European University, Budapest

The campaign that resulted in Britain voting to leave the EU was a populist appeal to the basest instincts of the economically, socially and culturally disenfranchised, fuelled by misinformation and a dismissal of any sincere debate of the potentially cataclysmic consequences. It follows the rise of populist parties using similar tactics in the 2014 European parliamentary election, and the unstoppable appeal of Donald Trump. We need to improve our understanding of these convulsions.

Systematic research on populism is still in its infancy. Social scientists are divided even on the definition, and on how it is distinct from more established phenomena such as the radical right, nationalism and xenophobia. Country experts often have good contextual knowledge of relevant cases — from France's Front National on the right, through Greece's Syriza on the left to the ideologically unidentifiable Five Star Movement in Italy. But scholars have only recently started to develop and collect high-quality data that are comparable across countries, such as the rhetoric of party leaders and the attitudes of their supporters. Such information will be crucial to understanding both the causes and the long-term consequences of populism.

In the short term, populist rhetoric may have the positive effect of engaging citizens in the democratic process. But these citizens rarely become informed. Responding to misleading emotional appeals, they can end up voting against their own interests — as with those in some UK regions who voted in droves to quit the union, despite being in receipt of the greatest EU largesse and having the smallest numbers of immigrants. In the longer term, populist politics can increase social polarization, erode trust and be profoundly damaging to society or the economy, as happened in oil-rich Venezuela, which is now practically a failed state.

It is possible that better institutional design and more effective responses to populist rhetoric could mitigate the powerful negative effects of such movements. In the absence of truly comparable data, we cannot yet assess what the most appropriate governance or communication strategies are to counter these politics. We do know what sort of research is needed.

It is crucial to develop measures that can be compared across countries. These should identify politicians who are populists, pinpoint which voters are susceptible and predict their responses to types of counter-messaging. Collection of such data is a resource-intensive task, but necessary to better understand the conditions under which populism rises, to chart whether its effect (even in moderation) is necessarily damaging, and if so, to work out how to defeat it.

In this way, we will be able to explore when and how past instances of populism — which have been studied for decades by Latin American scholars — are applicable today in Europe, the United States and beyond. Such studies will also determine to what extent one populist movement in Europe is analogous to another, even though they are often ideologically very different.

One promising effort is Team Populism (populism.byu.edu). This consortium of researchers, directed by political scientist Kirk Hawkins at Brigham Young University in Provo, Utah, is working to improve comparative assessments of populism. It brings together researchers from different levels of analysis (individuals, institutions, countries), various methodological convictions (researchers of mass attitudes, behavioural experimentalists, scholars studying politicians and discourse analysts), and different areas of expertise (scholars of the radical right in Europe and the populist left in Latin America). Such endeavours should help us to tackle the alarming kind of politics that led to Brexit.

PETER TURCHIN

Mine the past for patterns

Professor in the departments of ecology and evolutionary biology, anthropology and mathematics, University of Connecticut

Getting people to cooperate in very large groups such as the EU is difficult. Understanding how humans have been able to create cooperative societies is also hard, because we cannot readily run experiments. Nevertheless, much progress can be achieved by taking a scientific approach to analysing historical data.

As an example, the EU's rapid expansion from the original group of 6 states to the current 28 has clearly contributed to its dysfunction. Historians have a name for it: "imperial overstretch" (a classic example is the bloated empire of Alexander the Great, which fragmented soon after his death).

Dysfunction arises because, first, it's easier for six people (or six heads of state)to converge on a mutually agreeable course of action than for 28 to do so. Second, expansion beyond the western European 'core' brought together people, and politicians, from diverse cultures with different values, taking incompatible paths towards cooperation. Such normative and institutional mismatch created extra barriers to collective action.

Would European integration be better served by a more 'modular', step-wise approach? For example, central European countries already have their own 'integration nucleus' — the Visegrad Group of Poland, Hungary, Slovakia and the Czech Republic. Perhaps the EU would work better as a nested set of such groups rather than one large one that relies on informal arrangements between the more powerful states?

Such hypotheses need testing, empirically and systematically — with massive

historical databases that thoroughly sample the historical record (see, for example, the Seshat: Global History Databank; seshatdatabank.info). Here are some of the questions that need to be asked.

What administrative arrangements and political institutions aided cooperation in large empires (which often started as confederations), such as Rome, the Maratha Confederacy on the Indian subcontinent and the United States? What can we learn from the fate of the Habsburg Empire — the previous, failed, attempt at a 'European Union', formed through a series of dynastic marriages? Does gradual, incremental construction result in a longer-lasting union? What kind of hierarchy of political units works better: a flat one with a single level, or a nested, multilevel one?

This sort of analytical, predictive history — known as cliodynamics — can inform the design of a better, more cooperative EU (P. Turchin *Nature* **454,** 34–35; 2008). But there is a marked tendency among policymakers to deal with the economic and political crises of today as though they were unprecedented, leading them to repeat old mistakes.

The first step should be to invest massively in research on how human cooperation at large social scales has been achieved in the past. We all have a huge stake in European peace, prosperity and collaboration — even those outside Europe. We shouldn't just leave it to the politicians.

STEVEN J. BRAMS

Offer more voting options

Professor of politics at New York University

Of the many things that the United Kingdom's referendum has demonstrated, one is that voting, even by a plebiscite, need not be so crude. The stark choices offered to UK voters — remain in the EU or leave — belied the complexity of the issues involved. Instead, the British public should have been offered the choice of different levels of partial separation — three, say — bracketed by the two all-or-nothing options, giving them five choices in all.

True, more choices mean that a vote will be more finely divided, allowing one option to win with as little as 21% of the vote instead of the 52% that actually occurred. A system called approval voting obviates this problem, however, by allowing voters to indicate all courses of action they consider acceptable. The one with the greatest number of approvals wins.

The approach tends to find the most acceptable option over all, not the one with the strongest minority support that benefits from a divided field. Typically, the winner is a compromise, which receives support from voters on both sides of the ideological divide. It is one that most voters can live with, even if it is not their first choice.

Unlike voting systems in which voters can rank alternatives, which are used in such countries as Australia and Ireland, approval voting is a grading system — with only two grades (approve or don't approve). It avoids a serious problem with ranking systems in which there may be a cycle of top choices (majorities prefer A to B, B to C, and C to A); these cyclical majorities leave unclear which alternative should be the winner. Under approval voting, if A beats B, and B beats C, A beats C. So, except for when the result is tied, there is always an alternative that is approved more than any other.

"Voting, even by a plebiscite, need not be so crude."

Approval voting is now used to elect officers in several major scientific societies, including the American Mathematical Society and the Mathematical Association of America, each with tens of thousands of members. At my university, this system is used by the politics and economics departments to choose a chair. All tenured professors are candidates; the vote invariably elects candidates who bridge factions in the departments. In the US state of Oregon in 1990, approval voting was used in an advisory referendum on financing education. Voters could choose from five different levels of sales and income taxes, but because it was advisory, no choice was mandated.

If Britain's exit does precipitate referendums elsewhere in Europe, citizens should be given more than two options. This mimics the multiple choices that members of parliament typically have when they add and remove provisions from bills, promoting compromises. Shouldn't voters be afforded the same opportunities as their representatives?

SIMON GÄCHTER

Study how groups collaborate

Professor of the psychology of economic decision-making, University of Nottingham

Britain's decision to quit the EU can be interpreted as scepticism towards deep multilateral cooperation. This raises a fundamental question for social-science research: under what conditions can groups of people cooperate for the common benefit? So far, behavioural economics has overwhelmingly focused on investigating people's willingness to pull their weight for the benefit of their group — for example, how fishermen work together to secure the biggest catch for a village. But in a multilateral international cooperation, such as the EU, groups — or nations — work together to secure gains that exceed the possibilities of any individual nation going it alone. European research collaboration and the EU carbon market are prime examples.

Coalitions of diverse groups are necessary to address global challenges, but they are always jeopardized by the free-riding of groups who consider their self-interest — possibly fuelled by strong group identity — to be the only relevant criteria for collective decision-making. Witness the struggle to forge international climate agreements. Strong group identities can help people with a common sense of purpose to cooperate for example, in local communities, companies, universities, army units or sports teams. But such tribalism may undermine teamwork between groups, in particular when the costs of working together seem to be more salient than its benefits — as seems to be increasingly the case with citizens' views of European cooperation.

New experimental studies should in particular investigate the influence of identity on collaboration within and between groups when collective benefit and self-interest are in tension, or seem to be so.

Urgently needed are good data to validate pundits' hunches about people's motivations in voting in the referendum, and how those motives interacted with sociodemographics and social networks on and offline. For example, how important were immigration and identity? Which other considerations mattered for whom, where and why? There is impressionistic evidence about some motives but there are not enough hard data from large-scale representative surveys. These might also include simple behavioural experiments to probe the role of basic behavioural tendencies such as impatience, risk aversion, reciprocity, altruism and trust in strangers.

Answers to these questions are immediately important for evidence-based policy responses in the United Kingdom and in the EU. More broadly, understanding the relative importance of economic, sociological and psychological determinants of people's propensities to support transnational cooperation is crucial in an age of globalized challenges. ■

Mary Shelley, painted around 1840 by Richard Rothwell and housed in the National Portrait Gallery.

PHOTO: TARKER/BRIDGEMAN IMAGES

SCIENCE FICTION

The science that fed *Frankenstein*

Richard Holmes ponders the discoveries that inspired the young Mary Shelley to write her classic, 200 years ago.

In 1816, a teenager began to compose what many view as the first true work of science fiction — and unleashed one of the most subversive attacks on modern science ever written. Eighteen-year-old Mary Godwin (as she then was) had the idea for *Frankenstein, or The Modern Prometheus* that summer, while at the Villa Diodati on Lake Geneva in Switzerland, with her lover and future husband Percy Bysshe Shelley, and his friend and fellow poet Lord Byron. Forced inside by stormy weather, the group spent wild evenings telling ghost stories, while Byron's personal physician, the brilliant 20-year-old John William Polidori, regaled them with reports of the latest developments in medical science.

Mary's inventive mind was peculiarly primed to grapple with both literary and scientific controversy. Her mother was the feminist writer Mary Wollstonecraft, who had died from complications after Mary's birth. Her father was anarchist philosopher and novelist William Godwin, whose friends included chemists and pioneering electricity researchers Humphry Davy and William Nicholson, and the opium-addicted poet Samuel Taylor Coleridge. These influences shaped her youthful thinking, and were encouraged by Shelley, who had dabbled in science at the University of Oxford before being thrown out for atheism.

GOTHIC DRAMA

The myth of Victor Frankenstein, the crazed but idealistic young scientist who unwittingly lets loose his monstrous creation and struggles to accept responsibility, is a heady cocktail of gothic melodrama and disturbing speculation. It has proved astonishingly adaptable. The first theatrical version, *Presumption: or the Fate of Frankenstein*, opened at the English Opera House in London in 1823, to huge audiences and scandalous publicity ("Do not take your wives, do not take your daughters, do not take your families"). Mary Shelley attended, noting that "in the early performances all the ladies fainted and hubbub ensued!" There have been more than 90 dramatizations since, including the Danny Boyle-directed 2011 production at London's National Theatre, which opened with the Creature dropping naked from a huge, pulsating artificial womb. The story has also been adapted for more than 70 films, including James Whale's iconic 1931 *Frankenstein* starring Boris Karloff. In May this year, a Frankenstein ballet was staged at the Royal Opera House in London. Choreographer Liam Scarlett shrewdly analysed it as a love story: "The Creature is like an infant. He's desperately seeking a parent or loved one to take him through the world."

Although the myth is well known, the original novel is not. There are three versions. Mary Shelley began to write the first, probably as a short story, in two notebooks at Villa Diodati, expanding it during the winter of 1816–17 in simple direct prose of great intensity (the notebooks remained unpublished until 2008). The second, lightly edited by her husband and more literary in manner, was published in 1818. The third was radically revised by Mary Shelley alone, and was published in 1831, with a fascinating new introduction by her.

With each version, the basic plot remains the same, but the tone grows darker. Frankenstein becomes more passionate and ambitious, his science becomes more sinister and misdirected ("I felt as if my soul were grappling with a palpable enemy") and his Creature becomes more alienated and agonized. The 1831 introduction also contains an inventive, retrospective account of the storytelling competition at the villa. Mary now calls the book her "hideous progeny", and claims that the whole idea came to her instantly, like an emotional bolt of summer lightning on waking from a terrible nightmare. "I saw — with shut eyes but acute mental vision — I saw the pale student of unhallowed arts kneeling beside the thing he had put together. I saw the hideous phantasm of a man stretched out, and then, on the working of some powerful engine, show signs of life, and stir with an uneasy, half vital motion."

"The early chapters evoke the mysteries of experiment, naive excitement about electrical kites and the fascination of air pumps."

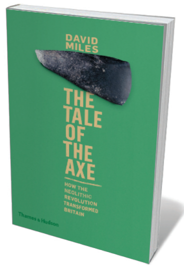
The book may, however, have had a more intellectual genesis. The best contemporary account of the ghost-story competition is Polidori's. A medical graduate of the University of Edinburgh, he had written his doctoral thesis on sleepwalking. Before the trip, he was commissioned by the publisher John Murray to keep a secret journal of Byron's adventures, and in this he recorded the villa party's speculative conversations and reading of German gothic "horror tales". Above all, he noted their wide-ranging discussions of fundamental scientific principles, and whether the human body "was thought to be merely an instrument". As Polidori put it, their brains "whizzed".

SCIENCE FACT

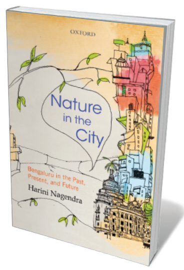
Polidori would have known about recent experiments in electrical resurrection techniques by Italian physicist Giovanni Aldini (nephew of bio-electrician Luigi Galvani), and the new anatomical theories of German physiologists such as Johann Friedrich Blumenbach. Also making waves were the fierce 'vitalist' debates at England's Royal College of Surgeons between John Abernethy and William Lawrence, about the possible existence of an electrical 'life-force' and the unique nature of human consciousness. These controversial ideas, alive in the great universities and research centres of Europe, fed into *Frankenstein*, and especially into the moral issues that it raised about the perils of scientific interference with nature.

Thus began a writing process involving careful research over many months. ▶

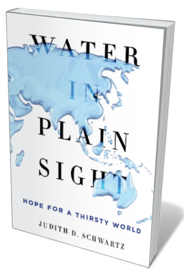
Books in brief



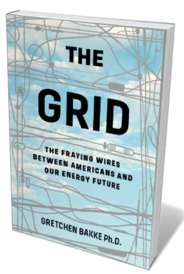
The Tale of the Axe: How the Neolithic Revolution Transformed Britain
David Miles THAMES & HUDSON *(2016)*
This illuminating treatise on the Neolithic era in Britain treats the polished-stone axe that gives the age its name as a portal into prehistory — a revelation of material, manufacture and function. Drawing on research riches from Turkey's Çatalhöyük site to Britain's Stonehenge, archaeologist David Miles contextualizes his core chronicle of how tools, farming and metallurgy arrived in the British Isles. As layered as the strata of an archaeological dig, this is a moving portrait of a people at a cultural and technological tipping point.



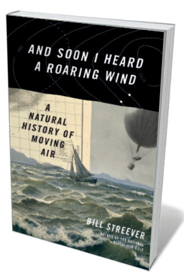
Nature in the City: Bengaluru in the Past, Present, and Future
Harini Nagendra OXFORD UNIVERSITY PRESS INDIA *(2016)*
With 10 million people and pell-mell development, Bengaluru (India's Silicon Valley, also known as Bangalore) is an old city in thoroughly modern flux. Urban ecologist Harini Nagendra's study looks at its deep ecological history, colonial role as India's garden city and current struggle with pollution, social exclusion and residents' increasing detachment from nature. Marshalling research from satellite imaging to interviews with slum dwellers, she concludes that "cities need to be ecologically as well as socially smart", and sees solutions in cross-city engagement of governance and civil society.



Water in Plain Sight: Hope for a Thirsty World
Judith D. Schwartz ST MARTIN'S *(2016)*
Water security demands holistic, ecosystem-oriented solutions, argues Judith Schwartz in this stellar global tour of innovative soil and biodiversity restoration and water harvesting. In Zimbabwe, ecologist Allan Savory reveals how intensified grazing by wild ruminants is enabling 95% of rainfall to soak into the soil, and rivers to recover. In Brazil, researcher Antonio Nobre exposes how deforestation damages the Amazon's unparallelled "forest-rain dynamics" and promotes drought. And in the Texas desert, permaculturalist Markus Ottmers unveils a built "ecosystem fuelled by variants of dew". Inspiring.



The Grid: The Fraying Wires Between Americans and Our Energy Future
Gretchen Bakke BLOOMSBURY *(2016)*
The US electricity grid, cultural anthropologist Gretchen Bakke reminds us in this cogent study, dominates US energy but is extremely vulnerable — and not just to gnawing squirrels. Nationalized and predicated on power plants, it's a poor fit with the variable, localized output of renewables. Bakke traces it inception by pioneers such as business magnate Samuel Insull through its technological, political and industrial evolution. Working towards a "self-healing, processor-dense 'intelligent' grid", she argues, is the key to energy resilience.



And Soon I Heard a Roaring Wind: A Natural History of Moving Air
Bill Streever LITTLE, BROWN *(2016)*
As his 2009 *Cold* and 2013 *Heat* (both Little, Brown) attest, biologist and nature writer Bill Streever is drawn to extremes. He now tackles strong winds, from cyclones to Santa Anas, for a scientific history of storms, meteorology and wind power, studded with pioneers such as seventeenth-century astronomer and trade-wind mapper Edmond Halley. A chronicle of Streever's voyage under sail from Texas to Guatemala is threaded through, giving a breezy immediacy to the story of how we learned to decode "moving air". **Barbara Kiser**

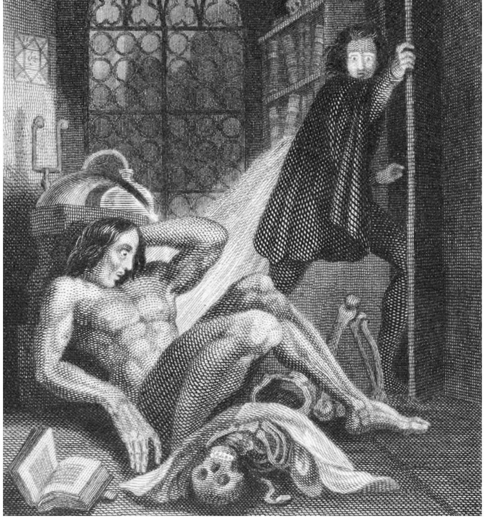
Shelley first mentions this in her journal for 24 July 1816. She was in Switzerland while walking above Chamonix towards Mont Blanc, absorbing the bleak landscape of the Mer de Glace glacier that would later fill the book's central confrontation between scientist and Creature. "Nothing can be more desolate than the ascent of this mountain ... we arrived wet to the skin ... I write my story". Her notes on triumphantly completing the first draft, "Transcribe and correct F[rankenstein] ... Finish transcribing" do not appear until April and May 1817, just four months before the birth of her third child, Clara. It is no accident that metaphors of pregnancy, birthing and parentage suffuse this novel about the creation of life.

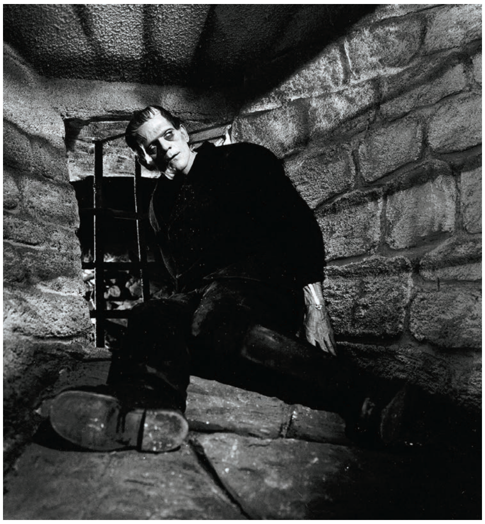
STREAMS OF INFLUENCE

In the intervening period of composition, back in England, Mary Shelley's journal reveals an impressive reading list. She absorbed the extreme accounts of polar exploration in George Anson's 1748 *Voyage Round the World*; the distinction between alchemy and chemistry in Davy's 1812 *Elements of Chemical Philosophy* (based on his famous London lectures); and the new concepts of brain development explored in Lawrence's physiological lectures, given in 1816–17. In Coleridge's 1798 poem *Rime of the Ancient Mariner*, she encountered the psychology of guilt and abandonment; in John Milton's 1667 *Paradise Lost*, the theme of the demonic outcast. Her husband also made clear, in his anonymous preface to the 1818 edition, that they had discussed the scientific poetry of Erasmus Darwin, in *The Temple of Nature, or The Origin of Society* (1803). Everything she devoured was brilliantly recast as a new genre: science fiction.

Thus, Davy's lectures at London's Royal Institution were subtly transposed, sometimes almost phrase by phrase, into those of the fictional Dr Waldman, praising the work of contemporary scientists to young Frankenstein. "These philosophers ... penetrate into the recesses of nature, and show how she works in her hiding places. They ascend into the heavens; they have discovered how the blood circulates, and the nature of the air we breathe. They have acquired new and almost unlimited powers; they can command the thunders of heaven, mimic the earthquake, and even mock the invisible world with its own shadows."

From her first draft, Mary had devised a complex structure that nests three autobiographical narratives one within the other like Russian dolls, each bringing a different interpretation to the Frankenstein myth. The first, often overlooked in adaptations, is by polar explorer Robert Walton. Told in the form of letters to his sister, it bookends the novel in the Arctic Ocean, and presents a moral enigma. Is the idealistic young Frankenstein essentially philanthropic, blindly ambitious or simply insane? And is his Creature evil or innocent — an ugly outcast or a persecuted victim longing for love?





Frankenstein's monster in the book's 1831 edition; played by Boris Karloff in 1931; and in a 2016 ballet.

The second autobiography is Frankenstein's own, particularly his thrilling discovery of the deep "enticements of science". These early chapters are among the first fictional presentations of the education of a young scientist, evoking the mysteries of experiment, naive excitement about electrical kites and the fascination of air pumps. Brilliantly transformed in the 1831 edition, these become more sophisticated references to galvanism, the necessity of mathematics, the genius of Isaac Newton and the intoxicating delights and dangers of charismatic science lecturing.

The third narrative, dramatically held back until halfway through, is the Creature's. Written in a wholly different stylistic register, it swings violently between desperate exclamations, poignant appeals and furious menacings. In the great showdown with Frankenstein on the Mer de Glace, the Creature begs the scientist to delve further into experimentation to create a female companion whom he can love.

Faced with this terrible ethical dilemma, Frankenstein agrees: this second creation scene, in a secret laboratory on the Orkney Islands off northeast Scotland, is also often overlooked. Fearful of the consequences, he destroys his female creation at the last moment, turning the disappointed Creature into a vengeful demon. So emerges the central drama of the novel. It is not merely the creation of life itself, the technical ambition of science, that is called into question. It is the unfolding moral choices and unforeseen ethical responsibilities that may come with scientific advances: artificial intelligence or artificial life, nuclear power or nuclear weaponry, the genome sequence or invasive genetic editing.

One added irony makes Shelley's novel much greater than any film — and greater indeed than its popular interpretation as an anti-science myth. It is that in these exchanges, paradoxically, the Creature becomes even more expressive and human than Frankenstein. He produces arias of speech, begging for justice, understanding, compassion and human rights. In the encounter in the Alps, the Creature declares himself Frankenstein's unique responsibility: "I ought to be thy Adam, but I am rather the fallen angel, whom thou drivest from joy for no misdeed ... Every where I see bliss, from which I alone am irrevocably excluded ... Misery made me a fiend. Make me happy, and I shall again be virtuous."

That is the enduring youthful genius and imaginative generosity of Mary Shelley's *Frankenstein*. It proclaims that the alien, the outcast, the rejected, finally must have claims on our humanity. And claims on our science, too. ■

Richard Holmes *is the author of* The Age of Wonder, *which won the 2009 Royal Society Prize for Science Books.*
e-mail: richard.holmes.biog@gmail.com

TOP: ILLUSTRATION: THEODORE VON HOLST/PHOTO: THE BRITISH LIBRARY BOARD, 11660.E.39.; MIDDLE: UNIVERSAL/THE KOBAL COLLECTION; BOTTOM: BILL COOPER/ROH 2016

Correspondence

Fresh strategies to save orangutans

The Bornean orangutan (*Pongo pygmaeus*) was listed as critically endangered by the International Union for Conservation of Nature this month, despite decades of conservation efforts. We urgently need fresh strategies to counteract habitat loss and hunting, and to mitigate the impacts of climate change.

Both species of Asia's only great ape (*P. pygmaeus* and the Sumatran orangutan, *P. abelii*) now face 'an extremely high risk of extinction'. Safeguarding protected areas and enforcing existing bans on orangutan killing are still important, but such efforts are insufficient in isolation (E. Meijaard *et al. Ann. N. Y. Acad. Sci.* **1249,** 29–44; 2012).

We need to maximize the conservation value of degraded lands and use an integrated, landscape-scale approach to reconcile economic development and biodiversity conservation. And we must promote sustainable management practices that are backed by rigorous, transparent certification mechanisms and that incorporate explicit economic valuation of the ecosystem services provided by orangutans and their habitats.

Crucially, success will also demand direct engagement and collaboration between conservationists and the mining, forestry and oil-palm industries.

Andrew J. Marshall *University of Michigan, Ann Arbor, USA.*
Serge Wich *Liverpool John Moores University, UK.*
Marc Ancrenaz *Borneo Futures, Kota Kinabalu, Sabah, Malaysia.*
ajmarsha@umich.edu

Vet reproducibility of biology preprints

Posting preprints in online repositories is common practice in the physical sciences and mathematics. It has been less satisfactory for preprints in the biological sciences, perhaps because the general standard of quality is inferior. Addressing the credibility of these submissions could bring free preprint servers such as bioRxiv (founded in 2013; bioRxiv.org) more into line with arXiv (arxiv.org), which has been running successfully for 25 years.

Introducing a staging process for submission could rectify the credibility problem. This would establish priority for new work and allow time for it to earn a seal of approval — thereby helping to speed publication after formal submission to a journal.

Initially, an abstract would be assessed by the editors of the preprint server to confirm suitability for posting. Submission of a short form of the paper would follow, with added details of raw data, materials and methods; this would allow other researchers to confirm, refute or comment on the results. After 6 months, the authors could submit a revised preprint in response to this feedback, together with evidence to support the paper's credibility and reproducibility; alternatively, they could withdraw the manuscript with dignity.

Roy Calne *University of Cambridge, UK.*
ryc1000@medschl.cam.ac.uk

SDGs: diseases that neglect no goals

The mapping of interactions between the United Nations' Sustainable Development Goals (SDGs) should guide efforts to leave no one behind (see M. Nilsson *et al. Nature* **534,** 320–322; 2016).

Take neglected tropical diseases (NTDs; go.nature.com/29te2hu) and the SDG target to end epidemics by 2030. Doing so will have a positive interaction with all SDGs, including those directed at poverty, hunger, education, work, growth and inequality (C. Fitzpatrick and D. Engels *Intl Health* **8** (S1), i15–i18; 2016). The example set by the largest-ever ongoing donation of medicines from drug firms (see go.nature.com/2ag1gnf) will help to realize the public–private partnership goal for health.

Community-led distribution of those medicines to more than 1 billion at-risk people is reinforcing other goals, such as women's empowerment, logistics infrastructure and non-discrimination against disability. Curbing mosquito-borne NTDs contributes to the goals of urban sustainability and resilience to climate change. Safe use of insecticides supports the goal of sustainable ecosystems. Controlling water- and animal-related NTDs is facilitating the goals of small-scale fishing and sustainable hydroelectricity and biofuels.

The World Health Organization is monitoring some of the most important positive interactions. For example, we are using NTDs to highlight inequalities, to target investment and to track progress towards universal access to water, sanitation and hygiene (see go.nature.com/29tpeph).

Christopher Fitzpatrick, Mathieu Bangert, Dirk Engels *WHO, Geneva, Switzerland.*
fitzpatrickc@who.int

SDGs: create a coordinating body

The United Nations' Sustainable Development Goals (SDGs) seem to be a disjointed hodge-podge of aspirations. I suggest that a 'panel of panels' would help to guide policymakers. This would collate the flow of information from existing UN panels to highlight connections between the SDGs (see also M. Nilsson *et al. Nature* **534,** 320–322; 2016).

This meta-panel could be modelled on international science platforms that coordinate multilateral environmental governance regimes, such as those for climate and biodiversity (see also P. M. Haas and C. Stevens in *Governing the Air* (eds R. Lidskog and G. Sundqvist) 125–161; MIT Press, 2011). It would similarly comprise independent experts from various countries and disciplines, drawn from academia, civil society and the private sector to capture different perspectives on sustainability.

The panel should be properly funded and not depend on a single funding source. It would meet regularly and report directly to governments through the UN High Level Political Forum. Its findings would need to be especially clear in areas of consensus and controversy.

Peter M. Haas *University of Massachusetts Amherst, USA.*
haas@polsci.umass.edu

A very preventable mammal extinction

James Watson blames climate change for the extinction of the Bramble Cay melomys, or mosaic-tailed rat (*Melomys rubicola*; *Nature* **534,** 437; 2016). In our view, simple management interventions could and should have saved this rodent.

With a population of at most a few hundred and a range limited to a single, small island (less than 5 hectares) that is susceptible to inundation in storm surges, the melomys was already living on the brink of extinction. Yet not an iota of effort was made, or resources committed, to reduce its extreme vulnerability.

Perhaps because the species lacked phylogenetic, cultural or ecological significance, responsible agencies seemed to treat it as inconsequential. A recovery plan developed in 2008 did not include plausible options for captive breeding or for translocation to nearby islands, and was never implemented.

Neglect and inaction caused this extinction — climate change was one of many threats that could have delivered the final blow. Such disregard must not be allowed to drive more extinctions.

John C. Z. Woinarski* *Threatened Species Recovery Hub, Australia.*
john.woinarski@cdu.edu.au
**On behalf of 4 correspondents (see go.nature.com/2aelgaj for full list).*

For News & Views online, go to
nature.com/newsandviews

HUMAN PERCEPTION

Amazon music

The people of a tribe called the Tsimane', who have been isolated from Western music, perceive music differently from Western listeners, raising questions about whether musical preference is innate or cultural. SEE LETTER P. 547

ROBERT ZATORRE

People have been making music for at least 35,000 years[1]. The remarkable diversity of musical expression across cultures raises the question of how the brain perceives musical sounds. Deep in the Bolivian Amazon lives a group of people, the Tsimane', whose music is largely isolated from Western influences. On page 547, McDermott *et al.*[2] report that the Tsimane' perceive harmony — the relationship between multiple simultaneous tones — differently from Western listeners.

Concepts of 'consonance' and 'dissonance' in Western music are based on frequency relationships between simultaneous tones. Periodically vibrating sources such as the human vocal cords give rise to sound at a fundamental frequency, produced by oscillation of the whole source, and at integer multiples of that frequency, corresponding to vibrations of half the length of the source, one-third its length and so on. This physical principle has long been thought to be a determinant of how listeners perceive harmony. For instance, tones whose frequencies are in simple integer ratios to one another, such as 3:2 (called the perfect fifth) or 5:4 (the major third), are considered consonant and more pleasant than dissonant ones such as 16:15 (the minor second). If this hypothesis is correct, then all listeners, regardless of their culture, should prefer similar tone combinations.

However, the authors found that the Tsimane' express no preference for consonance over dissonance. This result was replicated using several different stimuli, even dissonant versions of Tsimane' songs. In a clever control, the experimenters showed that the Tsimane' do prefer pleasant vocalizations, such as laughter, to unpleasant gasps, indicating that they understood what was being asked of them.

On the basis of their findings, McDermott and colleagues conclude that preference for harmonically related tones is not an innate feature of the human auditory nervous system, but is instead determined by exposure to music that contains harmonic relationships. Such a hypothesis provides much food for thought, but can be interpreted in several ways. For instance, one remarkable — and unusual — feature of Tsimane' music is that only one line is played or sung at a time (Fig. 1). Despite repeated attempts to encourage Tsimane' musicians to play or sing together, they demurred. It therefore seems likely that the lack of preference among the Tsimane' for harmonically related tones is linked to the fact that simultaneous tones are irrelevant in their musical culture.

Figure 1 | A Tsimane' man makes music. McDermott *et al.*[2] report that a Bolivian tribe of people called the Tsimane' show no preference for consonance over dissonance. This might reflect their musical culture.

What is perhaps most surprising is the general lack of preference among the Tsimane' for any one combination of tones over any other. Although music that contains simultaneous dissonant lines is relatively rare across cultures[3], dissonance is often used to produce sonic effects or textures[4], to create tension followed by resolution, or sometimes just for its own sake (for example, in contemporary jazz). Thus, rather than mapping onto a simple pleasantness dimension, dissonance and consonance may be better thought of as ways to manipulate sound expressively, and thereby engender emotions or moods. Again, the absence of multiple lines in Tsimane' music probably explains why the Tsimane' do not express a preference either way.

Speech may provide a useful parallel. When certain sound contrasts are absent from a given language (for example, the distinction between 'r' and 'l' sounds in Japanese), speakers lose the ability to distinguish between these sounds after a certain age[5]. The phenomenon represents tuning of the auditory system to the needs of the environment. Could something similar be happening here? We do not know how well the Tsimane' can distinguish consonant from dissonant tone combinations. The authors did show that the Tsimane' could determine (albeit with some difficulty) whether one or two tones were present on the basis of mistuning of the harmonics; but that does not quite answer the question. Lack of exposure to simultaneous tones may shape the auditory system such that the interactions between harmonics that influence preference in Western listeners are not processed in the same way, or perhaps simply not attended to.

Research into brain plasticity shows that neuronal responses to frequency and other features depend on the sounds that an individual is exposed to early in life[6]. Sensitivity to particular features may therefore change depending on the musical environment. Plasticity is also

relevant in current models of musical pleasure, in which cultural influences and experiences shape the brain's cortical circuitry, which in turn influences the brain's reward system[7].

This explanation is consistent with a strong role for environmental effects, in line with the authors' interpretation of their data. But does this idea rule out the possibility of innate factors? Not necessarily, because, despite their lack of preference for harmonically related tones, the Tsimane' did display a similar dislike to Western listeners for roughness — the sensation of sound that is elicited by tones that are close together in frequency, especially minor and major seconds. This finding suggests that there are probably some innate biological constraints on which environmental input operates. Indeed, auditory cortical neurons in macaques show differential responses to minor and major seconds compared with other intervals[8]; it would be hard to argue that this effect is mediated by the monkeys' musical culture.

The findings also raise the question of how or whether harmonic preferences arise when a culture develops the use of simultaneous sounds. Although the Western harmonic system is definitely not the only way that multiple sounds can be organized, its principles are not arbitrary, and allow for the complex relationships known as tonality. Moreover, melodic (sequential) frequency relationships can also be based in part on harmonic relationships, perhaps owing to exposure to the harmonics of voiced speech[9]. Tsimane' songs do use musical scales, as do most cultures[3]. McDermott *et al.* also mention that the Tsimane' use major seconds, minor thirds and perfect fourths, all of which are based on simple integer ratios, and which are also prevalent in Western music. So the influence of harmonic relations might be expressed in melody, if not in harmony. Simple integer ratios are also prevalent in rhythm across cultures[3] — it would be interesting to see whether this feature is preserved in Tsimane' music.

Although many questions remain, this work represents an important contribution to our understanding of how the diversity of human cultural expression can influence perception. More generally, it provides clues to how the environment interacts with the nervous system to produce all manner of complex behaviours, feelings and thoughts. ■

Robert Zatorre *is at the Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada.*
e-mail: robert.zatorre@mcgill.ca

1. Conard, N. J., Malina, M. & Münzel, S. C. *Nature* **460,** 737–740 (2009).
2. McDermott, J. H., Schultz, A. F., Undurraga, E. A. & Godoy, R. A. *Nature* **535,** 547–550 (2016).
3. Savage, P. E., Brown, S., Sakai, E. & Currie, T. E. *Proc. Natl Acad. Sci. USA* **112,** 8987–8992 (2015).
4. Vassilakis, P. N. *Sel. Rep. Ethnomusicol.* **12,** 119–144 (2005).
5. Werker, J. F. & Hensch, T. K. *Annu. Rev. Psychol.* **66,** 173–196 (2015).
6. Schreiner, C. E. & Polley, D. B. *Curr. Opin. Neurobiol.* **24,** 143–156 (2001).
7. Zatorre, R. J. & Salimpoor, V. N. *Proc. Natl Acad. Sci. USA* **110** (Suppl. 2), 10430–10437 (2013).
8. Fishman, Y. I. *et al. J. Neurophysiol.* **86,** 2761–2788 (2001).
9. Gill, K. Z. & Purves, D. *PLoS ONE* **4,** e8144 (2009).

This article was published online on 13 July 2016.

CLIMATE SCIENCE

Unexpected fix for ocean models

Computational models persistently underestimate strong currents that redistribute ocean heat. This problem is solved in models in which ocean eddies are damped by coupling of the atmosphere with the sea. **SEE LETTER P.533**

KATHRYN A. KELLY & LUANNE THOMPSON

In all ocean basins, strong currents are trapped at the western boundary as a consequence of Earth's rotation. The currents in the Northern Hemisphere carry warm subtropical water north along the coast to latitudes of about 35–40° N, then turn east to form boundary-current extensions. Much of the heat carried by these currents is released to the atmosphere[1], and the remainder provides heat to the ocean at higher latitudes. The eastward currents simulated by ocean models are persistently weaker than those observed. On page 533, Ma *et al.*[2] show how coupling the ocean to the atmosphere in models can strengthen the eastward currents, which should in turn improve climate predictions.

In simulations, a weak eastward current causes ocean-temperature biases, so that water near the coast is warmer than in the real world and water in the ocean interior is too cold. A weak current will also have a weak front (the temperature difference across the current is small). This effect has implications for the atmosphere, because observational analysis[3] suggests that strong fronts anchor mid-latitude storm tracks, whereas weak fronts do not. Modelling[4] that couples the ocean to the atmosphere at high atmospheric resolution (approximately 50 kilometres) has demonstrated the crucial role of ocean fronts in driving the atmosphere both in the boundary layer — the region in which flow is directly affected by Earth's surface — and deep into the troposphere, the lowest part of the atmosphere, where most weather phenomena occur. Ma *et al.* focus instead on how those atmospheric modifications feed back to the ocean.

Weak currents in simulations can be partly fixed by increasing the spatial resolution of ocean models from approximately 100 km to about 10 km. The higher resolution increases the ocean's kinetic energy in the regions near the currents to values comparable to estimates based on observations[3]. However, the distribution of energy in the ocean-only models differs from that observed: in observations, more energy is contained in the core of the current extensions than in the eddies that it sheds, whereas the reverse is true for the models[5].

Coupling an ocean model to an atmospheric model would seem likely to further weaken current extensions, rather than to strengthen them, by cooling the warm region south of the current and warming the cold region north of the current. Surprisingly, Ma *et al.* show that such coupling can instead strengthen the current in high-resolution models.

Focusing on the Kuroshio Extension, a boundary-current extension in the North Pacific Ocean, the authors compare coupled models that resolve ocean-eddy-scale (about 10 km) air–sea heat exchange with models that do not (Fig. 1). They show that ocean eddies drain energy from the current extensions, and that coupling of an eddy-resolving atmosphere model weakens the eddies by removing heat from warm ocean eddies and injecting heat into cold ones. An implication of this study is that the stronger current simulated by coupled models will deliver heat away from the western boundary into the ocean interior, as observed, and that this will improve climate prediction.

There are some limitations and further important implications of Ma and co-workers' work. The effect of coupling to a high-resolution atmospheric model on western boundary currents was demonstrated only for the Kuroshio Extension region. Although western boundary currents all have much in common, such coupling will not necessarily improve simulations in all basins. Moreover, the improvement comes at a steep price in terms of computer resources and time. The successful result also implies that ocean-only models will continue to simulate weak currents, because they lack the atmosphere–ocean feedback that lowers the energy of ocean eddies[6,7].

Accurate simulation of the transport and uptake of ocean heat and of atmosphere–ocean heat exchange is crucial for climate prediction.

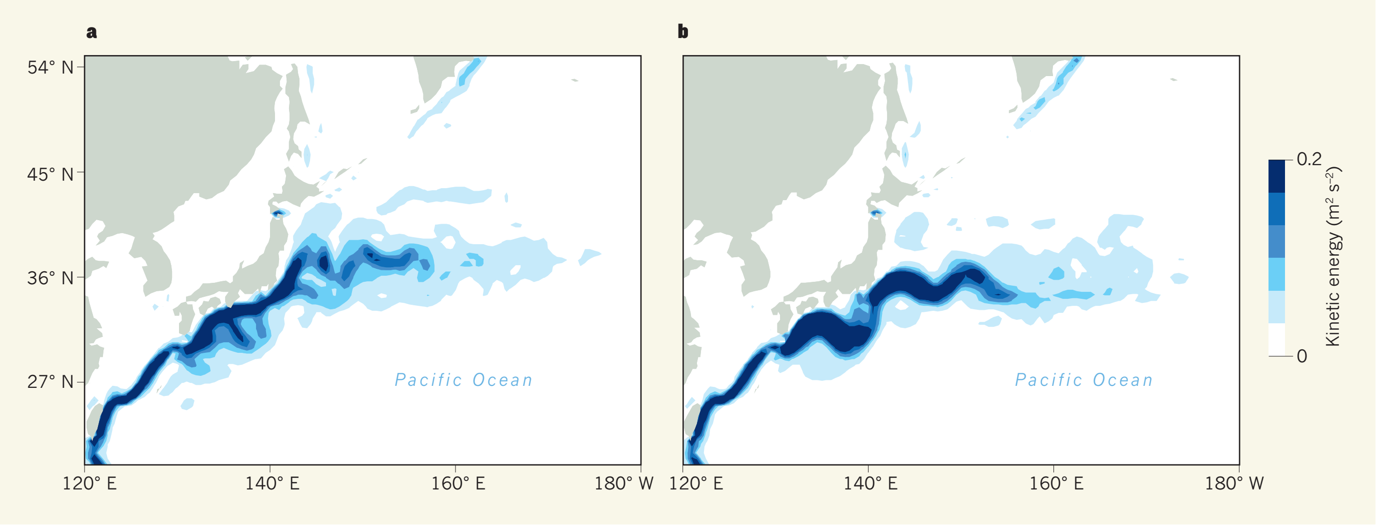


Figure 1 | Improved modelling of a strong ocean current. Ma *et al.*[2] modelled the Kuroshio Extension, a strong, eastward-flowing current in the North Pacific Ocean, using two different versions of a computational model. **a**, In the first version, small-scale (10 km) interactions between the ocean and the atmosphere were suppressed by having the atmosphere interact with only a 'smoothed' version of the ocean. A plot of the simulated kinetic energy of the Kuroshio Extension (shown in units of square metres per seconds squared) suggests a relatively weak, meandering current. Land is shown in grey. **b**, When the authors used a version that allows small-scale ocean–atmosphere interactions, the Kuroshio Extension is stronger and meanders less than in **a**, and has shifted southward, more closely resembling the observed current. These findings help to explain why ocean-only models persistently underestimate the strength of such currents.

Most models for century-scale and decadal prediction in the next assessment of the Intergovernmental Panel on Climate Change are being run at coarse resolution (100 km in the ocean); none will be at sufficiently high resolution to resolve eddies. Ma and colleagues have isolated a specific process that is missing in the lower-resolution climate models, and which leads to poor representation of ocean currents and therefore unrealistic air–sea heat exchange.

Boundary currents make a large contribution to heat transport by the ocean from the tropics to the high latitudes, which contributes to mid-latitude climate variations, storm-track steering and Arctic ice melt[8]. Accurate representation of these strong currents, such as the Gulf Stream, is thought to be necessary for climate prediction both in the near term (1–20 years)[9] and on the century scale. Ma and co-workers' study provides a strong and specific argument for increasing the spatial resolution of climate models to improve these predictions. ■

Kathryn A. Kelly *is in the Applied Physics Laboratory, Air-Sea Interaction and Remote Sensing Department, University of Washington, Seattle, Washington 98105-6698, USA.* **LuAnne Thompson** *is at the School of Oceanography, University of Washington, Seattle, Washington 98195-5351, USA.*
e-mails: kellyapl@uw.edu; luanne@uw.edu

1. Trenberth, K. E. & Caron, J. M. *J. Climate* **14,** 3433–3443 (2001).
2. Ma, X. *et al. Nature* **535,** 533–537 (2016).
3. Nakamura, H., Sampe, T., Tanimoto, Y. & Shimpo, A. In *Earth's Climate* (eds Wang, C., Xie, S. P. & Carton, J. A.) (Am. Geophys. Union, 2004).
4. Bryan, F. O. *et al. J. Climate* **23,** 6277–6291 (2010).
5. Maltrud, M. E. & McClean, J. L. *Ocean Modelling* **8,** 31–54 (2005).
6. Nonaka, M. *et al. J. Climate* **19,** 1970–1989 (2006).
7. Bryan, F. O., Hecht, M. W. & Smith, R. D. *Ocean Modelling* **16,** 141–159 (2007).
8. Holland, M. M., Bitz, C. M. & Tremblay, B. *Geophys. Res. Lett.*, **33,** L23503 (2006).
9. Meehl, G. A. *et al. Bull. Am. Meteorol. Soc.* **95,** 243–267 (2014).

GENETICS

Mitochondrial DNA in evolution and disease

Cellular organelles called mitochondria contain their own DNA. The discovery that variation in mitochondrial DNA alters physiology and lifespan in mice has implications for evolutionary biology and the origins of disease. SEE LETTER P.561

DOUGLAS C. WALLACE

The maternally inherited DNA found in cytoplasmic organelles called mitochondria encodes the central proteins involved in energy production — the main function of this organelle. Yet it has been assumed that the extraordinarily high sequence variability of mitochondrial DNA is of little consequence. On page 561, Latorre-Pellicer *et al.*[1] dispel this erroneous notion.

The authors transferred mitochondrial DNA (mtDNA) from a mouse strain called NZB to the nuclear DNA (nDNA) background of another strain, C57BL/6, and then compared C57BL/6 mice that harboured NZB or C57BL/6 mtDNA. The two mtDNA sequences differ in genetic variants that confer 12 amino-acid substitutions and 12 changes in RNA molecules involved in mitochondrial protein synthesis. Comparison of the mice throughout their lives revealed huge differences in mitochondrial function, insulin signalling, obesity and longevity. This and related studies[2,3] clearly demonstrate that naturally occurring mtDNA variation is not neutral, and that the interaction between mtDNA sequence variants and nDNA can have profound effects on mammalian biology.

Why should this be of general interest? It turns out that the amount of variation between NZB and C57BL/6 mtDNAs is about the same as that between two unrelated human mtDNAs, so mtDNA variation and its effect on nDNA gene expression is also relevant to people.

Mouse and human mtDNA sequences can

evolve only by sequentially accumulating mutations along radiating maternal lineages. For humans, functional mutations arose as women migrated out of Africa to colonize the rest of the world, modifying their cellular energy metabolism and allowing our ancestors to adapt to new regional environmental challenges. The mtDNA types (haplotypes) that acquired these environmentally advantageous mutations became widespread in their respective environments to give rise to regional groups of related haplotypes, called haplogroups. This regional selection explains why, of all the mtDNA lineages that evolved in Africa over the first 100,000 years of human history, only two mtDNAs (dubbed M and N) successfully left Africa 65,000 years ago to colonize the rest of the world. It also explains why only N mtDNAs colonized Europe, whereas both M and N colonized Asia, and why only five mtDNAs colonized the Americas (reviewed in ref. 4). Because human mtDNA diversity evolved from a single mtDNA, functional mtDNA variation that allows regional population isolation may also contribute to speciation[4]. Hence, mtDNA haplogroups are fundamental to the biology of both mice and men.

Because mitochondria have a bioenergetic role, it makes sense that mtDNA variation affects our physiology and our ability to adapt to environmental change. Variation in mtDNA genes can permit accommodation to new diets or adjustment to thermal stress and activity demands, and can even alter the regulation of cell death[4]. The correlation between human or mouse mtDNA variation and a broad range of traits, including longevity, physical capacity and, in humans, predisposition to a wide spectrum of metabolic and degenerative diseases and forms of cancer[4,5], confirms the functional importance of mtDNA variation.

Nuclear gene expression is also affected by mtDNA variation, owing to the role of the mitochondrial energy-production system in modulating the levels of high-energy molecules generated through mitochondrial metabolism. High-energy mitochondrial metabolic products, such as the molecules ATP, acetylCoA and α-ketoglutarate, drive the modification of cytoplasmic signalling proteins and also add molecular modifications to nuclear proteins, which, together with nDNA modifications, constitute the epigenome. Changes in cellular signalling and the epigenome regulate nDNA gene expression. This coupling between the mitochondrion and the nucleus is crucial because no cellular function can proceed without sufficient energy. The nucleus must ‘know’ that mitochondria can generate the required energy before proceeding with DNA replication and transcription, for example.

Studying cells that have the same nucleus but different levels of a pathogenic mtDNA mutation — a change in an RNA molecule in which nucleotide 3243 is guanine (3243G) instead of the normal adenine — has helped to define the nature of mtDNA–nDNA interactions. Each cell contains hundreds of mtDNA copies, so both mutant and normal mtDNAs can be present in the cell in different proportions, a state known as heteroplasmy. When the 3243G mutation is present at a frequency of 10–30%, patients can manifest diabetes or, in rare cases, autism; at 50–90%, the mutation manifests as neurological, heart and muscle problems; and at 100%, it can result in childhood disease and death. A study[6] of nuclear gene expression in cell lines containing different percentages of the 3243G mutation revealed that each of these clinical classes of heteroplasmy is associated with a distinct nuclear gene-expression profile. In humans, then, as in Latorre-Pellicer and colleagues’ mice, subtle changes in mitochondrial function resulting from mtDNA variation can have profound effects on nuclear gene expression, cellular physiology and individual health.

Uniparental inheritance of mtDNA is almost universal among animals, and mtDNA lineages are functionally different, so it might be predicted that artificially mixing two different mtDNA haplotypes in the same cell could be deleterious. Consistent with this prediction, mixing NZB mtDNAs with mtDNAs from a strain called 129 (whose mtDNA is similar to that of C57BL/6) in C57BL/6 mice results in mice that are hypo-active, hyper-excitable and have severe long-term-memory defects[7]. Hence, biparental inheritance of different mtDNAs can be deleterious, and the NZB and 129–C57BL/6 mtDNAs are functionally different.

Because mtDNA haplogroup variation has been shaped by environmental selection, it follows that mitochondria might be key sensors of environmental changes. Thanks to our growing understanding of the many regulatory roles of mitochondria, a holistic picture of cell biology has started to emerge (Fig. 1). In this scenario, changes in the environment would affect mitochondrial bioenergetics and alter the production of high-energy mitochondrial molecules. Altered concentrations of these mitochondrial molecules drive the chemical modification of cytoplasmic-signalling and epigenomic proteins, reprogramming nDNA gene expression. The altered nDNA gene-expression status then feeds back to modify mitochondrial gene expression and re-establish energetic homeostasis and health. However, if the local environmental changes are too severe to be managed by an individual’s mtDNA-haplogroup-defined physiological state, or if recent deleterious mtDNA mutations are too pathogenic, this feedback homeostasis system may fail, resulting in progressive energetic decline, disease and ultimately death.

Latorre-Pellicer and colleagues’ study provides direct evidence that naturally occurring mtDNA variation is fundamental to an individual’s characteristics and health. This information supports a model whereby mitochondrial physiological states determined by mtDNA variation sense environmental perturbations and send the appropriate

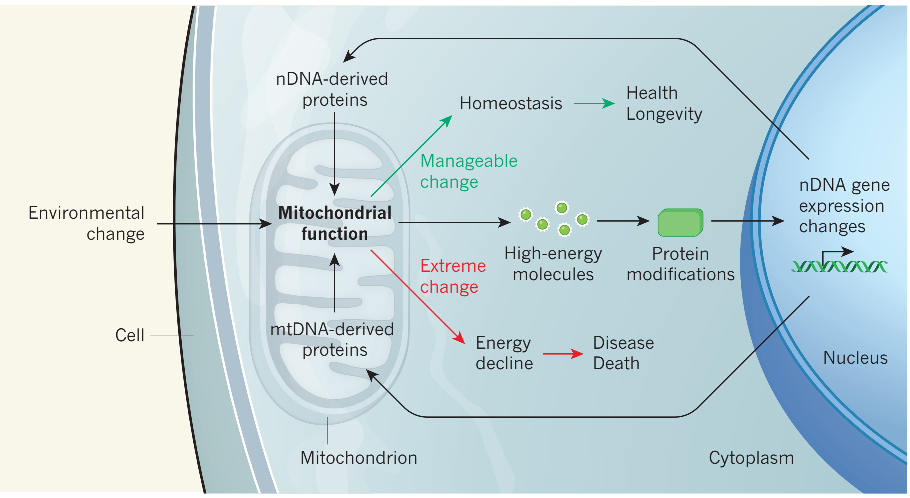


Figure 1 | Mitochondria as the central environmental sensor. Latorre-Pellicer *et al.*[1] report that the transfer of mitochondrial DNA (mtDNA) from one mouse strain to another has pronounced effects on biology, demonstrating that mitochondrial genetic variation is not neutral and that mitochondrial–nuclear interactions are of central importance to mammalian physiology. Mitochondrial function is directly influenced by environmental changes, so the mitochondrion must have a central role in mediating between environmental perturbations and genomic responses. High-energy molecules produced by mitochondria modify the cytoplasmic signalling proteins and ‘epigenomic’ proteins that regulate nuclear DNA (nDNA) expression. These changes reprogram gene expression, altering expression of nDNA- and mtDNA-derived proteins that act in and on the mitochondria — these alterations feed back on mitochondrial function. If energetic homeostasis can be re-established, health and longevity are preserved. However, if genetic or environmental changes are too extreme, energy production declines, leading to disease and even death.

signals to the nucleus to produce the optimal gene-expression response. ■

Douglas C. Wallace *is in the Center for Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA, and in the Department of Pathology and Laboratory Medicine, University of Pennsylvania.*
e-mail: wallaced1@email.chop.edu

1. Latorre-Pellicer, A. *et al. Nature* **535,** 561–565 (2016).
2. Yu, X. *et al. Genome Res.* **19,** 159–165 (2009).
3. Feeley, K. P. *et al. Cancer Res.* **75,** 4429–4436 (2015).
4. Wallace, D. C. *Cell* **163,** 33–38 (2015).
5. Picard, M. *et al. Proc. Natl Acad. Sci. USA* **112,** E6614–E6623 (2015).
6. Picard, M. *et al. Proc. Natl Acad. Sci. USA* **111,** E4033–E4042 (2014).
7. Sharpley, M. S. *et al. Cell* **151,** 333–343 (2012).

This article was published online on 6 July 2016.

SOLAR PHYSICS

Dynamo theory questioned

Observations of X-ray emission — a diagnostic tool for the mechanisms driving stellar magnetic fields — from four cool stars call into question accepted models of magnetic-field generation in the Sun and stars. **See Letter p.526**

PAUL CHARBONNEAU

Every star, including the Sun, hosts a magnetic field. One of the most notable products of stellar magnetic activity is X-ray emission (Fig. 1), most of which comes from active regions — areas on a star's surface where the magnetic field concentrates and whose best-known examples are sunspot groups. The commonly accepted theory of stellar magnetic-field generation is based on a dynamo process of electromagnetic induction: the mechanical energy of internal and surface plasma flows is converted into magnetic energy. According to this theory, less than one-tenth of 1% of the Sun's total luminosity is sufficient to drive the solar magnetic cycle, heat the corona, accelerate the solar wind and power all the eruptive phenomena that collectively make up solar activity.

But on page 526, Wright and Drake[1] show that the details of the dynamo theory still escape us. They report on the X-ray emissions from four stars detected by the Chandra and ROSAT space observatories. These stars are substantially colder than the Sun and have a different internal structure. If the chain of physical mechanisms that lead to X-ray emission in the Sun conformed to the dynamo theory, we would expect a different pattern of X-ray emission from these stars. Yet, Wright and Drake show that their X-ray emissions have the same pattern as the Sun's.

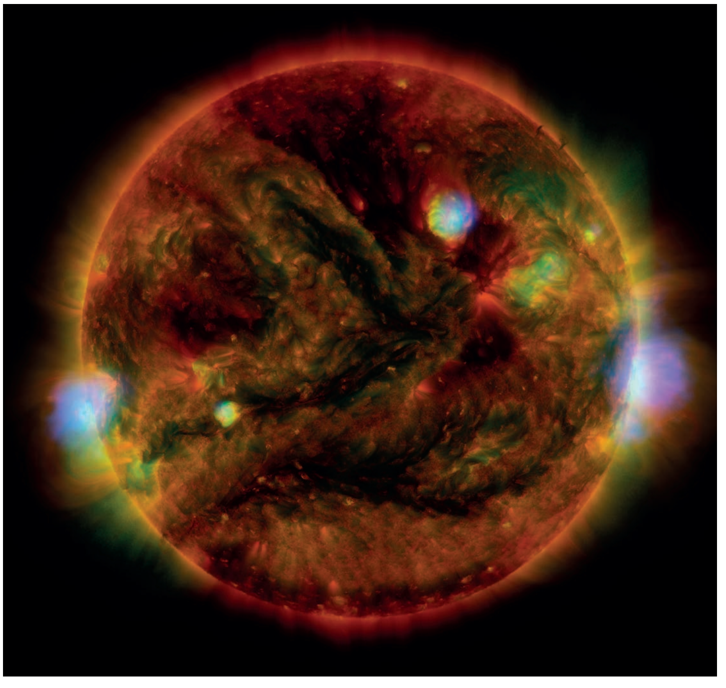
The solar-surface magnetic field can be observed in great detail because of the Sun's proximity to Earth. Such observations have revealed several spatiotemporal patterns that are driven by the dynamo. Perhaps most notably, polarity reversals occur with a regular 11-year period, and magnetically active regions are found to emerge ever closer to the solar equator as activity cycles unfold[2]. Conversely, observations of the activity of other stars are usually restricted to surface-averaged, global measurements that lack spatial information.

But, where stars are concerned, what researchers lack in detail is made up for in numbers. Surveys of X-ray emissions from large samples of stars — which, among other things, determine the dependence of the emissions on stellar mass, luminosity and rotation rate — provide information on global stellar activity and, in turn, insight into the underlying magnetic processes.

In Sun-like stars, X-ray emission increases with rotation rate up to a value of a few times the rotation period, and then levels off[3]. Information about the details of the internal dynamo is thus lost for fast-rotating stars, which fall within this plateau region of emission behaviour. Wright and Drake present an updated version of the emission-behaviour profile (see Fig. 1 of the paper[1]), based on data for hundreds of stars of varying masses and luminosities.

Within the solar dynamo theory, a key element of the complex causal chain that links the internal dynamo to surface X-ray emission is the tachocline, a transition region between the radiative core of the Sun and its convective outer layer. Solar plasma in the convective zone rotates at different rates, depending on latitude, whereas the radiative core rotates more or less as a solid body. This difference in rotational rate between the two zones produces a strong 'shear' stress in the tachocline, which helps to concentrate the diffuse magnetic field into structures called flux ropes. These magnetic flux ropes emerge at the surface and generate active regions[4].

Stars progressively less massive than the Sun have deeper convective envelopes, becoming fully convective at about 40% of the Sun's mass. Such stars no longer have a tachocline, so we should expect some qualitative change in their mode of dynamo action. This long-sought 'dynamo boundary' has not yet been detected. Fully convective stars have been found to emit the same level



NASA/JPL-CALTECH/GSFC/JAXA

Figure 1 | Short-wavelength emission from the Sun. In this composite image, extreme ultraviolet (red–yellow) was measured by the Solar Dynamics Observatory; low-energy X-rays (green) were measured by the Hinode spacecraft; and high-energy X-rays (blue) were measured by the Nuclear Spectroscopic Telescope Array. Wright and Drake[1] present observations of X-ray emissions from four stars that cannot be explained by currently accepted models of the solar dynamo.

of X-rays as solar-type stars that harbour a tachocline. However, all previously reported stars of this type are fast rotators, and therefore fall on the plateau region of the diagram that plots the relationship between X-ray emission and rotation rate (see the red circles in Fig. 1 of the paper[1]), meaning that the details of the internal dynamo action might well be lost.

Enter Wright and Drake. The authors have uncovered the X-ray emissions from four slowly rotating, fully convective stars, and all four fall on the slope part of the X-ray emission–behaviour profile. X-ray emissions therefore scale identically with the Rossby number (the ratio of convective flow speed to rotation rate) in stars with and without a tachocline.The authors thus argue that the tachocline cannot be an essential ingredient for stellar dynamo action, as it is in the currently accepted theory.

There are several possible ways to explain this quandary. Perhaps low-mass stars are not fully convective all the way down to their centres. Or maybe the pattern of stellar X-ray emission is dominated by a re-organization of the magnetic field that occurs in the stellar upper atmosphere[5], thus losing its 'memory' of the magnetic field's dynamo origin. And, of course, a tachocline might indeed be non-essential.

Numerical simulations[6,7] of global solar convection have demonstrated that solar-like large-scale magnetic fields undergoing regular polarity reversals can be produced wholly within a convection zone, without the need to extend the simulation down to the depth of the tachocline. Some of these simulations even generate rope-like structures of magnetic flux that rise to the top of the simulation domain in a solar-like manner[8,9]. Wright and Drake's results, together with such simulations, provide an impetus to rethink what we know about the solar and stellar dynamo. ■

Paul Charbonneau *is in the Département de Physique, Université de Montréal, Québec H3C 3J7, Canada.*
e-mail: paulchar@astro.umontreal.ca

1. Wright, N. J. & Drake, J. J. *Nature* **535,** 526–528 (2016).
2. Hathaway, D. H. *Living Rev. Sol. Phys.* **7,** 1 (2010).
3. Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K. & Vaughan, A. H. *Astrophys. J.* **279,** 763–777 (1984).
4. Fan, Y. *Living Rev. Sol. Phys.* **6,** 4 (2009).
5. Cheung, M. C. M., Rempel, M., Title, A. M. & Schüssler, M. *Astrophys. J.* **720,** 233–244 (2010).
6. Charbonneau, P. *Annu. Rev. Astron. Astrophys.* **52,** 251–290 (2014).
7. Hotta, H., Rempel, M. & Tokoyama, T. *Science* **351,** 1427–1430 (2016).
8. Nelson, N. J., Brown, B. P., Sacha Brun, A., Miesch, M. S. & Toomre, J. *Astrophys. J.* **739,** L38 (2011).
9. Fan, Y. & Fang, F. *Astrophys. J.* **789,** 35 (2014).

MICROBIOLOGY

Antibiotics right under our nose

Bacteria that are normally resident in the body have many roles in supporting health. Researchers have now identified a bacterial resident of the nose that produces an antibiotic that is active against a pathogen. SEE ARTICLE P.511

KIM LEWIS & PHILIP STRANDWITZ

There is immense clinical concern about the rise of antibiotic-resistant 'superbugs' — such as strains of the bacterium *Staphylococcus aureus* known as methicillin-resistant *S. aureus* (MRSA) that have developed resistance to several key antibiotics. Faced with the growth of resistant strains of bacteria, finding more antibiotics is an urgent necessity. Most antibiotics have been isolated from soil-living bacteria, but on page 511, Zipperer *et al.*[1] identify an antibiotic produced by a bacterial resident of the human nose that is active against strains of MRSA.

Staphylococcus aureus is an opportunistic human pathogen and, in addition to causing MRSA infections, is responsible for many infections, including those of the bloodstream (bacteraemia) and the lining of the heart (endocarditis), or infections surrounding prosthetic implants. *S. aureus* is found in the noses of around 30% of the population. This has prompted efforts to decolonize the nose with various antibiotics[2] — a controversial practice, considering efforts to reduce antibiotic use. But how is it that 70% of the population resist colonization by *S. aureus*?

The presence of several nasal bacteria, including other *Staphylococcus* species, is negatively correlated with the presence of *S. aureus*. But the mechanism that underlies this distribution pattern — and the presumed inhibition of *S. aureus* invasion (known as colonization resistance) — has been a mystery[3]. To investigate this potential bacterial antagonism, Zipperer and colleagues screened 90 *Staphylococcus* samples from the human nasal cavity, including many *Staphylococcus* species, to test for *S. aureus* growth inhibition.

One strain, *Staphylococcus lugdunensis,*

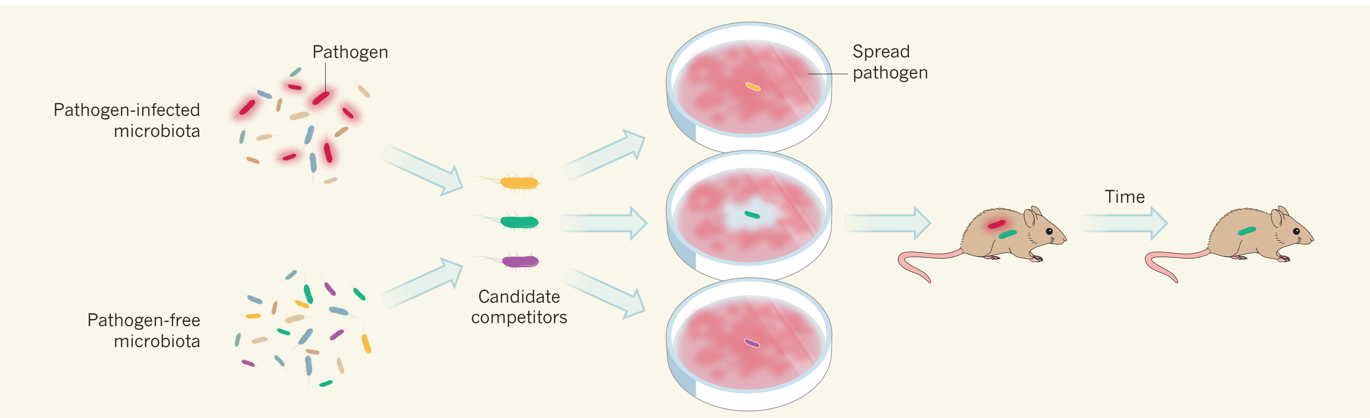


Figure 1 | An approach to identifying natural antibiotics. Zipperer *et al.*[1] used co-occurrence analysis of the human microbiota (the bacterial community in the body) to identify bacteria that do not co-occur with a target pathogenic bacterium. These candidate competitor bacterial strains were tested individually for the ability to inhibit growth of the pathogen, and those that did were then cultured and screened to see whether they conferred colonization resistance in an animal model. An antibiotic compound responsible for pathogen inhibition was isolated by the authors, and the compound, or the microbes that produce it, might be developed as therapeutics.

cleared a growing population of *S. aureus*, suggesting that the strain secretes a compound that causes the breakdown (lysis) of *S. aureus*. Using a library of *S. lugdunensis* mutants in which individual genes had been knocked out, the researchers identified a single mutant that was unable to inhibit *S. aureus*. Studying this mutant enabled identification of the compound responsible for *S. aureus* inhibition — a peptide antibiotic that the authors named lugdunin. Zipperer and colleagues' findings suggest that treatment with *S. lugdunensis* (or use of lugdunin) may be a valuable tool for the prevention of *S. aureus* colonization in the clinic.

It may seem surprising that a member of the human microbiota — the community of bacteria that inhabits the body — produces an antibiotic. However, the microbiota is composed of more than a thousand species[4], many of which compete for space and nutrients, and the selective pressure to eliminate bacterial neighbours is high. This intense competition can be observed by surveying patterns of bacterial communities in which the presence of a particular species is negatively associated with the presence of other species[5].

One probable mechanism for these negative interactions is the production of antibiotics. Although conventional antibiotic discovery has focused on mining compounds from soil-living bacteria, genome-wide analysis of the human microbiome has identified many gene clusters that encode enzymes associated with antibiotic production, such as polyketide synthases or non-ribosomal peptide synthetases. However, only a few such antibiotics have been characterized in the laboratory: strain-specific bacterial inhibitory molecules called bacteriocins[6,7]; fairly large peptide antibiotics called lantibiotics[8]; and the antibiotic lactocillin, which is active against several pathogenic bacteria[9]. Until now, evidence linking these compounds to the competitive advantage that they may confer has been lacking. The current study fills this gap.

Zipperer and colleagues found that lugdunin is active against several pathogens, including MRSA strains and *Enterococcus* bacteria that are resistant to the antibiotic vancomycin — and the bacteria treated with lugdunin did not develop resistance to it. In both *in vitro* and *in vivo* studies in mice, the ability of *S. lugdunensis* to outcompete *S. aureus* depends on the presence of a functional lugdunin biosynthesis pathway.

When the authors analysed 187 hospitalized patients, they found that *S. aureus* colonization was just 5.9% in individuals who carry *S. lugdunensis*, compared with 34.7% in individuals without *S. lugdunensis*. When produced by a bacterium that occupies a confined niche, it is clear that lugdunin and similar antibiotics have considerable power to influence bacterial community structure. Given that *S. lugdunensis* is present in only around 10% of the population and *S. aureus* is found in about 30% of the population, there are probably more antibiotics yet to be discovered that are responsible for *S. aureus* colonization resistance.

Lugdunin was found to be effective as a topical agent (applied to the surface) for treating an *S. aureus* skin infection in mice. Considering that lugdunin can inhibit the synthesis of major biopolymers (proteins, DNA and peptidoglycans) in *S. aureus*, it is probably a membrane-acting antibiotic, which would be challenging to develop into a systemic therapeutic because such compounds also tend to disrupt membranes of mammalian cells.

The method of discovery of lugdunin reported by Zipperer and colleagues provides a general approach to investigating antibiotic-driven colonization resistance of the human microbiota against pathogens (Fig. 1). By combining genomic and co-occurrence data, bacteria can be identified that are negatively correlated with the presence of a pathogen and that have the potential to produce antibiotics. These organisms, or the antibiotics they produce, might serve as drug-discovery leads. ■

Kim Lewis *and* **Philip Strandwitz** *are at the Antimicrobial Discovery Center, Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA.*
e-mails: k.lewis@neu.edu; p.strandwitz@northeastern.edu

1. Zipperer, A. *et al. Nature* **535,** 511–516 (2016).
2. Popoola, V. O. & Milstone, A. M. *J. Perinatol.* **34,** 805–810 (2014).
3. Liu, C. M. *et al. Sci. Adv.* **1,** e1400216 (2015).
4. The Human Microbiome Consortium. *Nature* **486,** 207–214 (2012).
5. Faust, K. *et al. PLoS Comput. Biol.* **8,** e1002606 (2012).
6. Kommineni, S. *et al. Nature* **526,** 719–722 (2015).
7. Zheng, J., Gänzle, M. G., Lin, X. B., Ruan, L. & Sun, M. *Environ. Microbiol.* **17,** 2133–2143 (2015).
8. Chikindas, M. L. *et al. Antimicrob. Agents Chemother.* **39,** 2656–2660 (1995).
9. Donia, M. S. *et al. Cell* **158,** 1402–1414 (2014).

In retrospect

Thirty-five years of endless cell potential

Pluripotent cells have the potential to differentiate into any cell type in the body. Their isolation and propagation from mouse embryos was pivotal for advances in understanding human development and disease.

M. AZIM SURANI

The human body consists of around 250 differentiated cell types, which all originate from about 15 cells in a structure called the inner cell mass that arises at the blastocyst stage of early embryonic development. These cells can each give rise to every cell type, but this pluripotent status is transient, with the cells soon differentiating into diverse types. This year marks the 35th anniversary of landmark studies by biologists Martin Evans and Matthew Kaufman[1], and Gail Martin[2], who independently captured and propagated these elusive pluripotent cells in a Petri dish for the first time, naming them embryonic stem (ES) cells. Studies of isolated ES cells have revolutionized our understanding of mammalian development, and enabled the simulation of human diseases *in vitro* (Fig. 1).

In the 1950s, it was reported[3] that mice, like humans, could develop testicular tumours called teratomas, which were derived from germ-cell precursors (precursors of eggs or, in this case, sperm) and contained various differentiated cell types not normally found in testes. The teratomas also contained embryonic carcinoma (EC) cells, which had key attributes of pluripotency.

However, EC cells could not efficiently differentiate into normal tissues if they were injected into blastocysts. In particular, their differentiation into sperm or eggs was not observed, meaning that EC cells could not be used to generate strains of mice that could pass experimentally induced mutations on to their offspring — a major goal of research into pluripotency. This shortcoming encouraged efforts to find other sources of pluripotent stem cells.

Evans and Kaufman established ES cells directly from mouse blastocysts that had been prevented from undergoing further development, for example by treating the mothers with steroids. Meanwhile, Martin grew inner cell masses from late-stage blastocysts using culture media in which EC cells had been grown previously, in the hope that the EC cells might have released factors conducive to the growth of ES cells. Both studies were successful, leading to the first isolations of blastocyst-derived ES cells that could be maintained indefinitely *in vitro*.

ES-cell isolation had an immediate and profound impact on the study of mammalian development. Unlike EC cells, ES cells could reliably give rise to germ cells[4],

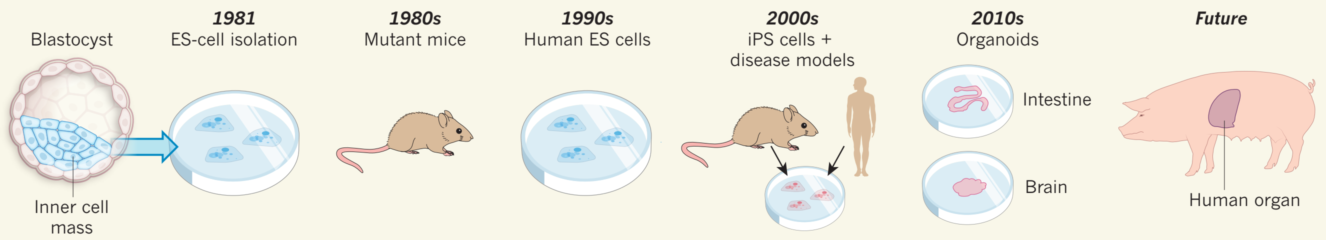


Figure 1 | Exploiting the potential of embryonic stem cells. In 1981, Evans and Kaufman[1] and Martin[2] independently isolated embryonic stem (ES) cells, which can give rise to all cell types, from the inner cell masses of mouse embryos at the blastocyst stage of development. During the 1980s, ES cells carrying experimentally introduced mutations were put back into blastocysts — some of the mutated cells developed into sperm and eggs, enabling the production of mutant mouse strains. In the late 1990s, human ES cells were first isolated. In the 2000s, study of ES cells from both species led to the direct reprogramming of adult cells into induced pluripotent stem (iPS) cells. iPS cells have since been used to simulate the progression of human diseases in culture. Since 2008, iPS and ES cells have also been used to generate synthetic *in vitro* structures called organoids, which share many of the properties of organs *in vivo*. In the future, ES cells may be used to grow human organs in host animals.

making it possible to introduce specific genetic mutations into mouse strains. To briefly outline the process, ES cells are genetically engineered to harbour a specific mutation, and are then transplanted into a host blastocyst. The resulting embryo is a chimaera — some of its cells will contain the mutation, others will not. If germ cells carry the mutation, it can be passed to the offspring. The consequences of the mutation can then be examined in detail.

A great many mutant mouse strains have been generated in this way, enabling detailed analyses of the functions of individual genes under physiological conditions. Genes vital for embryonic development, especially those that establish the body plan, were mutated to study their precise role[5]. Moreover, human diseases could be modelled in other mammals. For example, a gene that causes cystic fibrosis in humans was mutated in mouse ES cells[6], and mice carrying this mutant gene were then created to provide an animal model for the disease.

The development of large-scale mutation strategies such as gene traps (in which mutations are introduced into DNA at random through the insertion of a genetic construct that produces a 'reporter' protein) led to the production of mutation libraries — a useful resource of cells from which mice carrying particular mutations could be made. Before the isolation of ES cells, all such research was limited by the availability of naturally occurring mutations. But today, any gene can be investigated at will.

Following this impressive progress, attention shifted to humans. In 1998, human ES (hES) cells were eventually isolated[7]. Although not molecularly identical to mouse ES cells, hES cells nonetheless possess the key properties of pluripotency — self-renewal and an ability to differentiate into diverse lineages. Moreover, they provide opportunities for studying early human development.

Ethical considerations prevent hES cells from being used for germline therapy, in which genetic changes would be introduced into hES-derived germ cells with the aim of correcting faulty genes. However, two studies published last year[8,9] showed that germ-cell precursors can be derived from hES cells, paving the way for studies of human germ cells and indicating that it might one day be possible to make sperm and eggs in culture. Some key genes involved in the induction of human germ-cell precursors, however, are not involved in mouse germ cells[8,9]. Thus, these studies, with many others, have revealed key differences between mouse and human development.

"ES-cell isolation had an immediate and profound impact on the study of mammalian development."

Studies on ES cells also defined the four key regulatory genes that control pluripotency, which in 2006 led to a major advance — the conversion of differentiated cells into induced pluripotent stem (iPS) cells[10–12]. iPS cells have since been derived from patients and differentiated into specific cell types, to simulate the initiation and progression of disease in culture. Moreover, iPS cells are currently being used to develop *in vitro* models for testing potential therapeutic compounds to treat diseases such as Alzheimer's. Human ES cells and iPS cells might some day be used as part of cell-replacement therapies in diseased tissues, although the greatest advances in their use will initially probably be in basic research and in screens for therapeutic agents.

One useful feature of ES cells is their ability to self-organize into differentiated tissues and to assemble into structures that have recognizable attributes of organs. In the past decade, this quality has been exploited to produce synthetic organs called organoids in culture. So far, human organoids have been generated that model the brain, intestine, liver and kidney, among others[13–15]. More-advanced 3D models that approximate physiological settings are likely to be developed in the future using tissue-engineering methods. Attempts are now under way to generate human organs in animals, by incorporating hES cells into blastocysts from other species. This might alleviate the unmet demand for organs for transplantation.

Advances made following the isolation of pluripotent ES cells have coincided with progress in other technologies, such as live imaging, single-cell analysis and genome editing[16]. When combined to study human organoids, these approaches will allow analysis of the outcomes of increasingly sophisticated manipulations of genes and of the genetic elements that regulate their expression. Such studies will provide unprecedented understanding of how genes individually and collectively regulate human development, and how such regulation can go awry.

The history of ES cells shows how basic research has led to revolutionary advances. The future presents us with endless opportunities to explore how a single cell transforms into a multicellular adult organism. ■

M. Azim Surani *is at the Wellcome Trust/Cancer Research UK Gurdon Institute, Cambridge, CB2 1QN, UK.*
e-mail: a.surani@gurdon.cam.ac.uk

1. Evans, M. J. & Kaufman, M. H. *Nature* **292,** 154–156 (1981).
2. Martin, G. R. *Proc. Natl Acad. Sci. USA* **78,** 7634–7638 (1981).
3. Stevens, L. C. & Little, C. C. *Proc. Natl Acad. Sci. USA* **40,** 1080–1087 (1954).
4. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. *Nature* **309,** 255–256 (1984).
5. Dollé, P., Izpisúa-Belmonte, J. C., Brown, J., Tickle, C. & Duboule, D. *Prog. Clin. Biol. Res.* **383A,** 11–20 (1993).
6. Ratcliff, R. *et al. Nature Genet.* **4,** 35–41 (1993).
7. Thomson, J. A. *et al. Science* **282,** 1145–1147 (1998).
8. Irie, N. *et al. Cell* **160,** 253–268 (2015).
9. Sasaki, K. *et al. Cell Stem Cell* **17,** 178–194 (2015).
10. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
11. Takahashi, K. *et al. Cell* **131,** 861–872 (2007).
12. Yu, J. *et al. Science* **318,** 1917–1920 (2007).
13. Sasai, Y., Eiraku, M. & Suga, H. *Development* **139,** 4111–4121 (2012).
14. Lancaster, M. A. & Knoblich, J. A. *Science* **345,** 1247125 (2014).
15. Takasato, M. *et al. Nature* **526,** 564–568 (2015).
16. Doudna, J. A. & Charpentier, E. *Science* **346,** 1258096 (2014).

doi:10.1038/nature18942

Rapid signalling in distinct dopaminergic axons during locomotion and reward

M. W. Howe[1] & D. A. Dombeck[1]

Dopaminergic projection axons from the midbrain to the striatum are crucial for motor control, as their degeneration in Parkinson disease results in profound movement deficits. Paradoxically, most recording methods report rapid phasic dopamine signalling (~100-ms bursts) in response to unpredicted rewards, with little evidence for movement-related signalling. The leading model posits that phasic signalling in striatum-targeting dopamine neurons drives reward-based learning, whereas slow variations in firing (tens of seconds to minutes) in these same neurons bias animals towards or away from movement. However, current methods have provided little evidence to support or refute this model. Here, using new optical recording methods, we report the discovery of rapid phasic signalling in striatum-targeting dopaminergic axons that is associated with, and capable of triggering, locomotion in mice. Axons expressing these signals were largely distinct from those that responded to unexpected rewards. These results suggest that dopaminergic neuromodulation can differentially impact motor control and reward learning with sub-second precision, and indicate that both precise signal timing and neuronal subtype are important parameters to consider in the treatment of dopamine-related disorders.

Dopaminergic projections from the substantia nigra pars compacta (SNc) to the striatum are vital for motor control, and their degeneration in neurodegenerative disorders such as Parkinson disease results in severe motor problems[1–3]. However, little evidence for movement-related dopamine signalling has been observed; instead, most recording methods report that unpredicted rewards induce phasic signalling (hundreds-of-millisecond bursts), which are superimposed on slowly varying background activity[4–10]. The leading model to explain the dual role of dopamine in reward and movement posits that phasic bursts in striatum-targeting dopamine neurons act to drive reward-based learning, while slow variations in firing (tens of seconds to minutes) in these same neurons bias the system towards or away from movement[11–15]. However, there is little evidence for or against this model.

Most information about *in vivo* dopamine dynamics comes from extracellular single unit recordings in the ventral tegmental area (VTA) and SNc[6–8,16–21] (Supplementary Discussion). This approach provides functional characterization of individual neurons with single-spike temporal resolution, but it cannot discriminate neurons by their projection targets. Moreover, midbrain single unit recordings alone cannot reliably discriminate dopaminergic from non-dopaminergic neurons[8,22] and are blind to the local control exerted on dopamine terminals in the target regions[23,24]. Fast-scan cyclic voltammetry, on the other hand, measures changes in dopamine release directly in striatal terminal regions, but these measurements of extra-synaptic dopamine, averaged over hundreds of micrometres, are insensitive to synaptically confined dopamine and potentially heterogeneous signalling across different axons[25]. To measure both phasic (hundreds of millisecond) and slowly varying (tens of seconds to minutes) dopamine signalling in the striatum during reward delivery and animal movement, and to determine whether dopamine projections to the striatum broadcast homogenous or heterogeneous signals, we used two-photon calcium imaging and fibre photometry to establish the striatal dopamine signalling patterns occurring in locomoting mice receiving unpredicted rewards. We then tested the role that these activity patterns play in motor control by optogenetically stimulating striatum-targeting dopamine axons.

Locomotion related dopamine signalling

A new approach was established to image the activity of dopamine projection axons in dorsal striatum[26] (Fig. 1a–d). DAT-cre mice ($n=6$) were injected in the midbrain with adenovirus AAV1-Syn-flex-GCaMP6f[27], leading to specific expression of GCaMP6f in dopamine neurons widespread throughout SNc and VTA (Fig. 1d). A chronic imaging window was then implanted over dorsal striatum (Extended Data Fig. 1b) to enable two-photon imaging[26] (Methods, Extended Data Fig. 2i) of the GCaMP6f-expressing dopaminergic projection axons (Fig. 1c). Axons were clearly visible, but were densely packed (Fig. 1c, Extended Data Fig. 1a), presumably representing overlapping axonal arbors from numerous different dopamine neurons[28]. Mice were then head-restrained over a cylindrical treadmill, and time-series videos of the labelled axons were acquired (in the dark) during rest, self-initiated locomotion, and/or delivery of unexpected rewards (Fig. 1e).

Calcium influx into axons and terminals, caused by axonal action potentials[29,30] (with local modulation possible[23,24]), triggers neurotransmitter release[31–33], and therefore, changes in GCaMP6f fluorescence[27] provide an indirect measure of dopamine release in dorsal striatum[34]. These GCaMP6f fluorescence changes are well-suited for determining whether mouse locomotion (100 s of millisecond timescale accelerations) is associated with rapid phasic dopamine signalling (sub-second-duration GCaMP6f transients), slowly varying dopamine signalling (tens of seconds to minutes-duration transients or changes in baseline) or both (Methods, Extended Data Fig. 3a–c).

As a first metric to quantify dopamine signalling in dorsal striatum, we summed the fluorescence from all dopamine axons in each frame of each imaging field (60–250 μm diameter fields sampled at 28 Hz) to generate a mean fluorescence activity trace (mean change in fluorescence, $\Delta F/F$ trace; Fig. 1e, bottom). These traces revealed frequent transients during treadmill locomotion (Fig. 1e), and these varied in duration and amplitude but generally occurred on sub-second timescales (duration 678 ± 460 ms (mean ± s.d.); amplitude $126 \pm 67\%$ $\Delta F/F$; Extended Data Figs 2g, h, 3e). Longer duration transients often contained multiple peaks, as expected from summation of closely spaced, short duration transients[27] (Fig. 1e, Extended Data Fig. 3).

[1]Department of Neurobiology, Northwestern University, Evanston, Illinois 60208, USA.

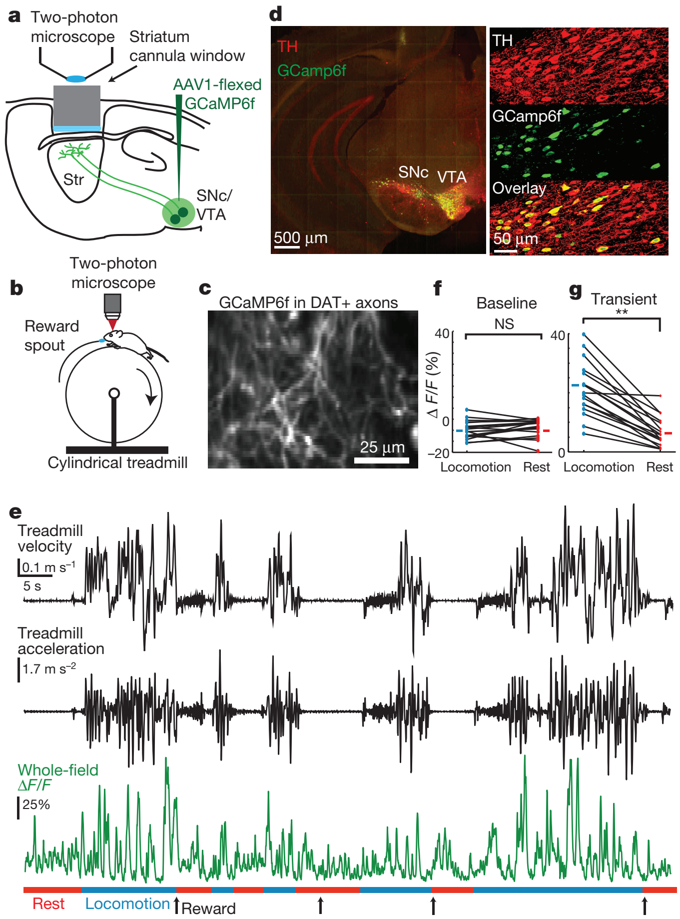


Figure 1 | Locomotion related signalling in dorsal-striatal-projecting dopamine axons. a, Schematic of methods. **b**, Mouse running on treadmill under microscope. **c**, Representative field of GCaMP6f-labelled dopamine axons in dorsal striatum. DAT, dopamine transporter. **d**, GCamp6f expression in dopamine neurons (green) overlaid with tyrosine hydroxylase (TH) immunofluorescence (red). Right insets are at high magnification. **e**, Average whole-field $\Delta F/F$ fluorescence from a representative field in one mouse (bottom, $n=6$ mice total) and corresponding treadmill velocity (top) and acceleration (middle) during locomotion. **f**, Comparison of mean baseline (periods with no significant calcium transients) whole-field $\Delta F/F$ between locomotion and resting periods. **g**, Comparison of mean whole-field fluorescence change (mean of significant calcium transients, excluded baseline periods) between locomotion and resting periods. **f**, **g**, Each point represents mean $\Delta F/F$ for running or resting over one session for each field ($n=17$ sessions from 6 mice, lines connect same field/session, short lines: mean over sessions); $^{**}P<10^{-5}$, Wilcoxon rank-sum test. NS, not significant ($P>0.05$).

Transients were typically synchronous across the densely innervated fields, presumably reflecting widespread co-activation of many projection axons (Extended Data Fig. 1), and were more frequent during spontaneous locomotion than resting (non-reward) periods (Fig. 1g, Extended Data Fig. 4h; $P=1.8\times10^{-4}$, Wilcoxon sign-rank test across all fields, $n=6$ mice). However, no significant difference in the fluorescence baseline (periods with no significant calcium transients) was found between locomotion and resting (Fig. 1f; $P=0.45$, Wilcoxon sign-rank test across all fields, $n=6$ mice). Moreover, the duration of nearly all significant dopamine transients ($n=2{,}087$) was far less than the duration of the average locomotion bout (mean locomotion bout 10.3 ± 11.2 s, Extended Data Fig. 3e). Thus in dorsal striatum, treadmill locomotion is associated with widespread and synchronous sub-second transient increases in GCaMP6f fluorescence in dopamine axons, suggestive of phasic dopamine release, but is not associated with detectable long-duration transients or sustained changes in baseline fluorescence levels, suggesting little slowly-varying dopamine release.

Dopamine signalling to acceleration bursts

We first examined whether the widespread fluorescence transients in dorsal striatum displayed a timing preference with respect to locomotion initiations on the treadmill (sharp rest–run transitions; Fig. 2a, d, f, Extended Data Fig. 4a, b). On average and at individual onsets, GCaMP6f transients increased rapidly just before (~100–150 ms) acceleration onsets occurring at locomotion initiation (Fig. 2a, Extended Data Fig. 4a, j, l, o; cross-correlation peak at -143 ± 250 ms, significantly different from 0 lag, $P=0.003$, Wilcoxon sign-rank test across peak correlation times, Fig. 2d, f). Thus, signalling across a widespread population of dopaminergic axons leads locomotion initiation in dorsal striatum, suggesting that sub-second dopamine signalling has a role in initiating locomotion bouts.

We then examined whether the widespread fluorescence transients displayed a timing preference with respect to discrete phases of continuous mouse locomotion movements. Movements consisted of rhythmic acceleration bursts on the treadmill over a broad frequency range centred on ~2.5 Hz (Extended Data Fig. 4e).

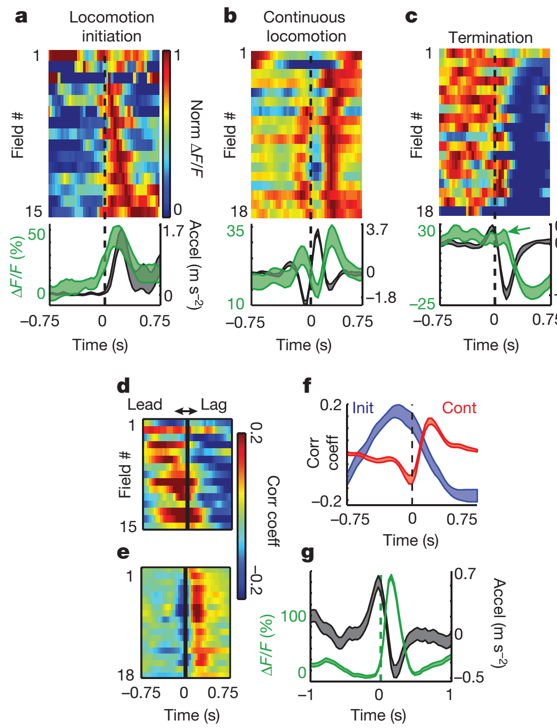


Figure 2 | Phasic dopamine signalling displays a sub-second timing preference with respect to acceleration bursts. a, Top, mean whole-field $\Delta F/F$ triggered on accelerations (accel) at locomotion initiations; each row is mean for each field/session ($\Delta F/F$ normalized for each row and sorted by peak time). Bottom, mean acceleration (black) and whole-field $\Delta F/F$ (green) all triggered on locomotion initiations (mean across all fields). **b**, Same as **a**, except triggered on accelerations during continuous locomotion periods. **c**, Same as **a**, except triggered on locomotion terminations; note lack of $\Delta F/F$ transient peak (arrow) after final acceleration. **d**, Correlations between acceleration and whole-field fluorescence at different relative time-lags (that is, cross-correlations) during locomotion initiation periods (sorted by peak correlation time); each row is mean for each field/session. Corr coeff, correlation coefficient. **e**, Same as **d**, but during continuous locomotion periods. **f**, Mean cross-correlations between acceleration and whole-field fluorescence for initiation (blue) and continuous (red) locomotion periods. **g**, Mean acceleration (black) triggered on all short duration calcium transients (green, mean of transients) during continuous locomotion across all fields. Shaded regions in **a**–**c**, **f**, **g** denotes mean ± s.e.m.; $n=6$ mice in **a**–**g**.

Indeed, fluorescence transients displayed a delayed timing preference with respect to acceleration bursts during continuous locomotion (Fig. 2b, e, f, Extended Data Fig. 4c, d), where on average they followed the onsets (zero crossing) of individual acceleration bursts (Fig. 2b,g; cross-correlation peak at 215 ± 90 ms, significantly different from 0 lag, $P = 1.66 \times 10^{-4}$, Wilcoxon sign-rank test across 18 fields; Fig. 2e, f, Extended Data Fig. 5a, b; $n = 6$ mice), an effect that was also clearly observed with individual isolated transients (Extended Data Fig. 4c, k). Importantly, the acceleration-associated calcium transients were observed before the mice ever received rewards on the treadmill (Extended Data Fig. 4h, i), and the signalling was highly similar after the introduction of unpredicted rewards. Thus, contrary to current widely accepted models, these results establish the existence of a widespread population of dopaminergic axons projecting to the dorsal striatum that display rapid phasic signalling associated with ongoing accelerations, which are expressed independently of reward expectation.

We then investigated whether the dopamine signalling during continuous locomotion might affect subsequent movements. We observed a significant correlation between calcium transient amplitudes and the amplitudes of the immediately following acceleration bursts (Spearman's Rho = 0.13, $P = 0.006$, all transient-acceleration pairs; binned data, Extended Data Fig. 5d), despite a weaker temporal association (compared to previous acceleration bursts, Extended Data Fig. 5a, b). Furthermore, just before locomotion termination (stopping), calcium transient amplitudes associated with the last acceleration burst (before termination) were reduced compared to the transient amplitudes expected from similar acceleration bursts during continuous locomotion (Fig. 2c, arrow). Thus, during continuous locomotion, dopamine axon transients seem to be more temporally associated with the previous (rather than following) acceleration burst (Fig. 2b, e, f, g, Extended Data Fig. 5a, b) and may positively impact the next acceleration burst (Extended Data Fig. 5e), perhaps acting as a feedback signal to promote plasticity or continued locomotion.

Locomotion triggered by dopamine axon stimulation

To test whether sub-second timescale signalling in dopamine axons can trigger or alter locomotion, channelrhodopsin[35] (ChR2) was expressed in midbrain dopamine neurons of DAT-cre mice ($n = 7$) and the dopamine axons in the dorsal striatum were (bi-laterally) optically stimulated (Fig. 3a, Methods). Optical stimulation consisted of 166 or 83 ms duration bursts of 8-ms pulses (pulses repeated at 60 Hz); these bursts were then repeated at 3 or 6 Hz, respectively (see Fig. 3b; frequencies corresponding to rhythmic accelerations measured during average (3 Hz) or fast (6 Hz) locomotion; Extended Data Fig. 4e) to create stimulation trains. This was intended to generate rapid phasic dopamine axon signalling similar to that observed during locomotion. The laser stimulation trains reliably trigged rapid transitions from resting to locomotion (or occasionally rocking forward and back; Fig. 3c–g; statistically significant acceleration increase averaged across all mice and sessions, $P < 0.01$, shuffle test; statistically significant increase in acceleration, averaged across all sessions for 6 out of 7 individual mice, $P < 0.01$, shuffle test), although not every stimulation led to locomotion (Fig. 3d, e). Transitions to locomotion were rapid: for 3 Hz burst trains, acceleration significantly increased from rest after the first burst (that is, ten 8-ms pulses; significantly increased acceleration compared to shuffle ~160 ms after burst onset, $P < 0.01$; Fig. 3g; similar to timing during spontaneous initiations, Fig. 2a, d, f, Extended Data Fig. 4l), whereas for 6 Hz burst trains, acceleration significantly increased from rest after two bursts (that is, two bursts of five 8-ms pulses; significantly increased acceleration compared to shuffle ~240 ms after burst onset, $P < 0.01$; Fig. 3g). Interestingly, 3 or 6 Hz stimulation trains that continued after locomotion initiation were capable of shifting the acceleration frequency in individual mice (Fig. 3h) towards the stimulation frequency during the ensuing locomotion bout. All mice ($n = 7$) displayed higher frequency accelerations during locomotion bouts initiated during 6 Hz stimulation than during 3 Hz stimulation (Fig. 3i, k, significant shift, $P < 0.01$, Wilcoxon rank-sum test) despite equivalent mean velocities during these periods (Fig. 3l). In two mice, acceleration power spectra peaks matched the respective stimulation frequencies (one of the two mice in Fig. 3h, top) and accelerations during locomotion were significantly entrained to the laser stimulations ($P < 0.01$, shuffle test; Extended Data Fig. 6a, b; entrainment also significant across all mice, $P < 0.01$ shuffle test, Extended Data Fig. 6c, d). Notably, the same optical stimulation applied to control mice without ChR2 expression failed to trigger locomotion or shift or entrain locomotion frequency ($n = 7$; Fig. 3f, j, k, Extended Data Fig. 6c, d). Thus, optogenetic stimulation of dorsal-striatum-projecting dopamine axons can rapidly trigger locomotion bouts and shift or entrain rhythmic acceleration frequency, demonstrating that signalling in dorsal striatal dopamine axons can impact locomotion movements on sub-second timescales.

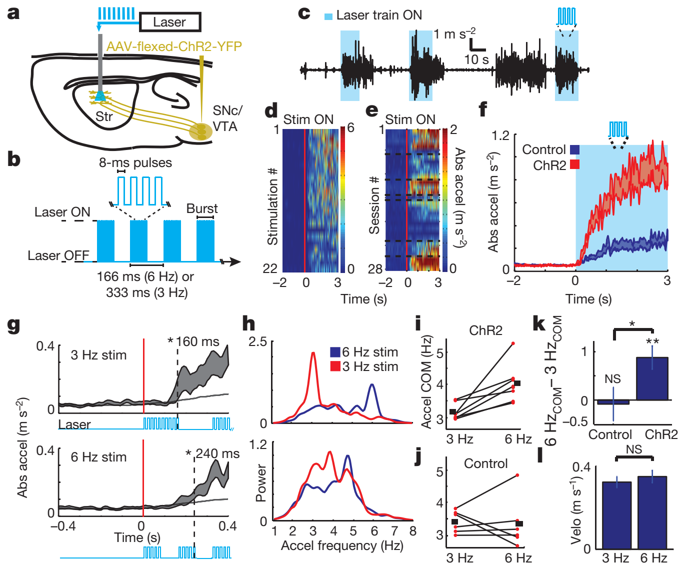


Figure 3 | Pulsed optogenetic stimulation of dorsal-striatum-projecting dopamine axons can rapidly initiate locomotion and control acceleration frequency. a, Schematic of methods. **b**, Pulsed laser delivery protocol for dopamine axon ChR2 stimulation (used for all data presented in this paper). **c**, Representative acceleration from a single stimulation session (blue regions, laser train stimulation to the dorsal striatum). **d**, Absolute value of mouse acceleration aligned on onset of all laser stimulation trains (red line) during a representative single session from one mouse (each row represents a single trial stimulation from rest). **e**, All mean accelerations aligned on laser stimulation trains applied to mice at rest ($n = 7$ mice total, each row represents mean over one session; rows corresponding to sessions for each mouse grouped together; sessions from different mice separated by dashed lines). **f**, Mean absolute (abs) value of mouse accelerations aligned on the onsets of laser stimulation trains applied to mice at rest (mean across all stimulation onsets, $n = 161$ and 267 onsets for ChR2 and control, respectively, in all sessions and mice, 3 and 6 Hz stimulation included). **g**, Same data as **f**, but stimulations separated into 3-or 6-Hz groups and zoomed-in time (blue, laser stimulation). Dashed lines, time when acceleration becomes significantly ($P < 0.01$) greater than random shuffle (thin grey line, mean random shuffle). **h**, Power spectra of acceleration for two representative mice during locomotion periods initiated during 3- and 6-Hz stimulations. **i**, **j**, Centre of mass of acceleration power spectra (COM) for each ChR2 (**i**) or control (**j**) mouse for locomotion periods initiated during 3-and 6-Hz stimulations. Horizontal bars indicate means, lines connect same mouse. **k**, Mean difference between centre of mass of acceleration power spectra during 3 and 6-Hz stimulations (differences calculated for each mouse) averaged across control or ChR2 mice. **l**, Mean velocity across mice during 3 and 6-Hz stimulations included in **k**. Shaded regions in **f** and **g** denote mean ± s.e.m. *$P < 0.05$; **$P < 0.01$ (compared to 0), Wilcoxon rank-sum test.

Dopamine axon functional heterogeneity

Unpredicted rewards are well-known to trigger phasic bursts of dopamine signalling[6–9,13,18]. Surprisingly, we observed little or no phasic response to randomly delivered water rewards in the mean whole-field $\Delta F/F$ traces; instead the signals decreased in accordance with the deceleration from locomotion to reward consumption or remained flat when reward was delivered at rest (Extended Data Fig. 4f, g). However, a small, sparse population of axons were found that signalled during reward (Fig. 4a). To characterize reward and locomotion signalling at the individual axon level, we then sparsely labelled small populations of midbrain dopamine neurons (Fig. 4b, Extended Data Fig. 7), resulting in sparse labelling of axons in dorsal striatum. This enabled us to resolve the fluorescence changes from putative single axons (Fig. 4c, d, Extended Data Fig. 8a, b) and revealed that, as suggested by the widespread labelling experiments (Fig. 4a), putative single axons in the same dorsal striatum region can differentially signal unpredicted reward and locomotion (Fig. 4d–i). Thus, the dopaminergic projection to the dorsal striatum is heterogeneous with respect to unpredicted reward and locomotion, but is dominated by locomotion signalling.

We then examined whether these functionally distinct axons emanate from different midbrain nuclei[36]. Injections were localized primarily to either SNc ($n = 5$ mice) or medial VTA ($n = 5$ mice) (Extended Data Fig. 7). In total, 73 isolated putative single axons (Methods) were imaged from SNc and 98 from VTA, of which 63 out of 73 (86%) and 72 out of 98 (73%), respectively, responded to locomotion and/or reward events ($P < 0.01$, Wilcoxon rank-sum test, locomotion versus rest; during unpredicted reward delivery, $P < 0.01$, shuffle test versus non-reward periods; or during both). The large majority of the responsive VTA axons signalled unpredicted reward (Fig. 4k, l, n; 53 out of 72 axons, 74% reward responsive). These responses were consistent with documented single-unit electrophysiology and fast-scan cyclic voltammetry[5,8–10,18,37]: they were absent on omission trials, scaled with reward-size, were not present to licking in the absence of reward, and were present whether reward was delivered during running or rest (Extended Data Fig. 9). In contrast to the VTA, such prominent reward signalling was absent from active dopamine axons emanating from SNc (Fig. 4k, l, n, Extended Data Fig. 9; 0 out of 63 axons). Instead, all event responsive axons from SNc displayed significantly increased signalling during locomotion compared to rest (Fig. 4n; 63 out of 63 axons). This signalling was transient rather than sustained (Fig. 4j, m, Extended Data Fig. 2g, h) and displayed a timing preference with respect to acceleration similar to the whole-field measurements (Fig. 2a–g), indicating that single SNc axons, rather than different subpopulations, can signal during locomotion initiation and continuous locomotion (Extended

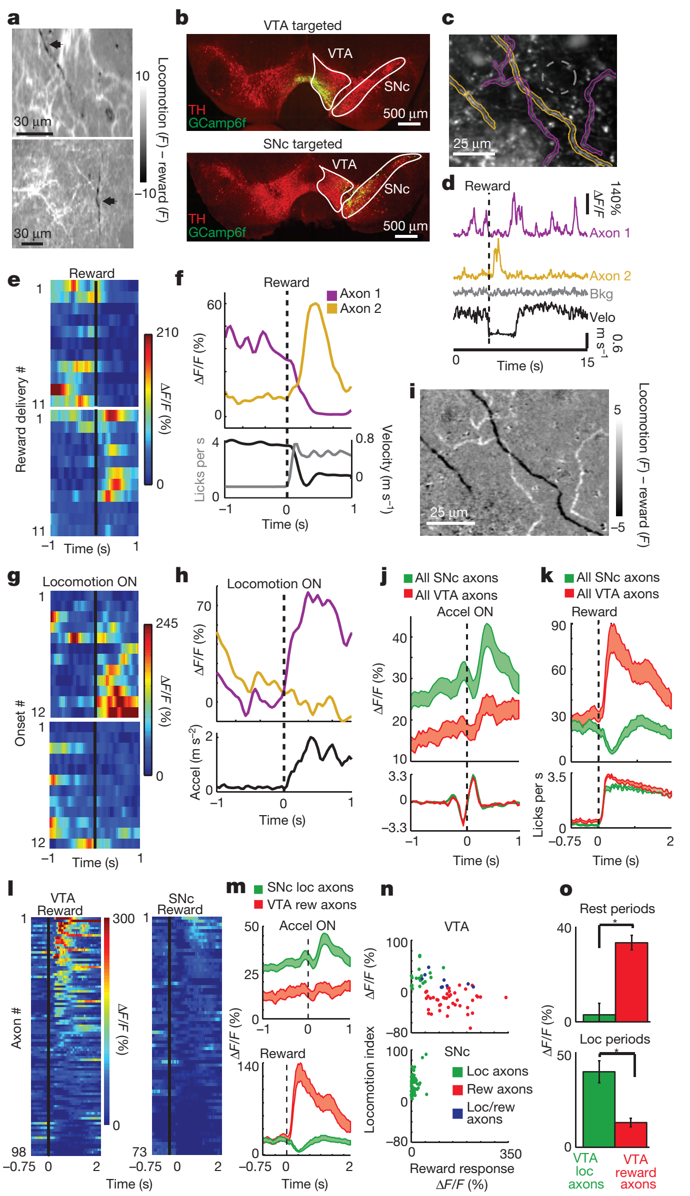


Figure 4 | Functional heterogeneity and anatomical origin of dorsal-striatum-projecting dopamine axons. a, Mean image during locomotion periods minus mean image during reward periods for two representative dopamine axon fields (arrows, axon segments signalling more during reward). **b**, Coronal sections showing restricted GCamp6f expression (green) in VTA (top) and SNc (bottom) overlaid with tyrosine hydroxylase immunofluorescence (red). **c**, Representative maximum fluorescence projection image from one field in a mouse with GCaMP6f expression localized to VTA ($n = 5$ total VTA-targeted mice). Gold and purple regions are identified regions of interest (ROIs) from putative single axons (grey, background). **d**, $\Delta F/F$ from ROIs in **c** aligned with treadmill velocity and reward time. Bkg, background. **e**, Reward-triggered $\Delta F/F$ from the purple (top) and gold (bottom) ROIs from **c**; each row represents a single reward delivery. **f**, Mean reward-triggered $\Delta F/F$ (gold, purple, top), velocity (black, bottom) and licking (grey, bottom) for the two axons in **c** over the recording session. **g**, **h**, Same as **e**, **f**, except triggered on locomotion onsets. **i**, Same as **a**, but for field shown in **c**. **j**, Mean ΔF/F triggered on acceleration onsets for all putative single axons from all SNc-injected ($n = 5$ mice, 73 axons, green) and VTA-injected ($n = 5$ mice, 98 axons, red) mice. Bottom, mean acceleration. **k**, Mean $\Delta F/F$ triggered on reward for all putative single axons in **j** from SNc (green) and VTA (red). Bottom, mean mouse licking triggered on reward delivery. **l**, Reward-triggered $\Delta F/F$ for all putative VTA (left) and SNc (right) axons (same axons in **j**, **k**, sorted by reward response magnitude); each row represents mean $\Delta F/F$ for each axon. **m**, Mean $\Delta F/F$ triggered on acceleration onsets (top) and rewards (bottom) for all significant ($P < 0.01$) VTA reward only responsive ($n = 43$ axons from 5 mice, red) and SNc locomotion responsive ($n = 63$ axons from 5 mice, green) axons. **n**, Reward response versus locomotion index for all VTA ($n = 5$ mice, top) and SNc ($n = 5$ mice, bottom) putative axons. Green, significant locomotion; red, significant reward; blue both significant; neither significant not shown. **o**, Mean $\Delta F/F$ during locomotion (bottom) and rest (top, non-reward periods during rest) periods for all locomotion ($n = 19$ axons, 4 mice green) and reward ($n = 43$ axons, 5 mice red) responsive VTA axons. $*P < 10^{-5}$, Wilcoxon rank-sum test. Shaded regions in **j**, **k** and **m** denote mean ± s.e.m.

Data Fig. 8c–e). Only a minority of the total responsive VTA (Fig. 4k, m–o; 29 out of 72 axons, 40%) and reward signalling VTA (Fig. 4m–o; 10 out of 53 axons, 19%) axons significantly increased signalling during locomotion compared to rest, but similar to the SNc projections, this signalling was transient and acceleration-associated. Overall, reward and locomotion signalling were negatively correlated in the heterogeneous VTA axon population (Fig. 4n; Rho = −0.51, $P = 5.7 \times 10^{-6}$). The presence of both reward and locomotion signalling axons from the VTA probably reflects some within-region heterogeneity, however, some expression spillover into SNc[36] or SNc/VTA border region may contribute to the locomotion population. Taken together, these results demonstrate a previously unknown functional heterogeneity within dorsal-striatum-projecting dopamine axons, which, for the most part, arise from different anatomical nuclei: sparsely projecting reward signalling axons originate primarily from VTA while densely projecting SNc originating axons generate an acceleration-associated locomotion-signal.

Dopamine signalling functional topography

Owing to the prominence of unpredicted reward signalling in previous recordings[4,7,9,17,19,37–39], the small fraction of reward signalling dopamine axons in dorsal striatum was surprising. However, because ventral striatum receives most of its dopaminergic innervation from VTA, whereas dorsal striatum receives most from SNc[36], we reasoned that locomotion and unpredicted reward signalling might be similarly graded across the striatum. Single-photon photometry[34,39] was used to record from the GCaMP6f-expressing dopamine axons along the dorsal-ventral axis (Fig. 5a; $n = 5$ mice). This method lacked single axon resolution, but provided a means of recording throughout the striatum with minimal tissue damage. Recordings from dorsal striatum revealed acceleration-associated dopamine signalling highly similar to that observed with two-photon imaging (Figs 1e, 2b versus Fig. 5b–d, Extended Data Fig. 10a–f; serving as a control for the cortical lesion required for two-photon imaging). More ventrally, acceleration-associated signalling declined while unpredicted reward signalling increased (locomotion index versus depth, Rho = 0.7, $P = 6 \times 10^{-6}$; reward signal versus depth: Rho = −0.65, $P = 5.2 \times 10^{-5}$; Fig. 5c, d, Extended Data Fig. 10g) and dominated in ventral striatum (Fig. 5c, d, Extended Data Fig. 10g). Consistent with this result, pulsed optogenetic stimulation of dopamine axons in ventral striatum had comparatively little effect on locomotion (Extended Data Fig. 6e–h; $n = 4$ mice, same ChR2-expressing mice as used in Fig. 3). No significant dopamine signalling was seen in the overlying cortex (Fig. 5c, d). Thus, a functional topography in dopamine signalling exists across the striatal dorsal–ventral axis[34], with locomotion and reward signalling dominating dorsal and ventral striatum, respectively.

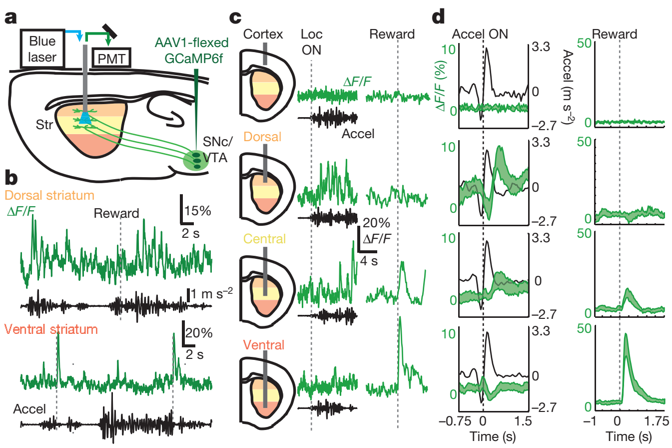


Figure 5 | Functional topography of reward and locomotion dopamine signalling across striatum dorsal-ventral axis. a, Schematic of photometry methods. **b**, Representative photometry $\Delta F/F$ (green) and treadmill accelerations (black) during locomotion and reward delivery (dashed lines) periods from dorsal (top) and ventral (bottom) striatum from same mouse (same session, $n = 5$ mice total). **c**, Representative photometry $\Delta F/F$ from individual locomotion initiations (left) and reward deliveries (right) measured at indicated depths (not averaged). **d**, Mean $\Delta F/F$ across all photometry recording sessions ($n = 5$ mice) triggered on acceleration onsets (left) and reward deliveries (right) at depths indicated in **c** (black, mean acceleration).

Discussion

The rapid, phasic, acceleration-associated locomotion signalling of dopamine projection axons described here represent a new mode of dopaminergic modulation of striatal circuitry and establish strong evidence for the long-elusive role for dopamine neurons in fast-timescale motor-control. This result is perhaps surprising given the little evidence for movement-related dopamine signalling previously observed; this difference may be due to task and/or methodological differences (Supplementary Discussion). Our results support new models of dopaminergic signalling that incorporate rapid, sub-second modulation of movement by dopamine[16,20,21] over current models that posit a permissive involvement by a slowly varying background[12,14,15,40] (Extended Data Fig. 10h–k). The phasic signalling occurred in the absence of action requirements or rewards, suggesting a general role in internally driven movement control (Extended Data Fig. 4).

Our optogenetic stimulation experiments support the idea that the phasic signalling in dopamine axons observed with optical recording can trigger and entrain locomotion movements on a sub-second timescale, demonstrating the utility of using measurements of natural neuronal dynamics[41] to define stimulation[35] parameters for effecting a desired behavioural outcome[42]. Dopamine signalling is part of a complex circuit regulating voluntary movement, and therefore is unlikely to instruct movement alone. The behavioural effects of these signals may therefore be strongly modulatory, rather than deterministic, an idea supported by the variability in signalling timing and stimulation effects on locomotion (Fig. 3d, h–k, Extended Data Fig. 4j, k), although this could also be partly owing to the limited region imaged or stimulated. Dopamine signalling could interact differently with the many converging components of the movement network during movement initiations versus continuous movements. At initiations (Fig. 2a, d, f, Extended Data Fig. 4a, b, j), dopamine may provide the necessary motivational drive to execute an intended movement. During ongoing movements, dopamine signalling lagging behind accelerations (Fig. 2b, e, f, g, Extended Data Fig. 4c, d) may modulate upcoming movements (Extended Data Fig. 5d, e), determine whether dorsal striatum circuitry maintains an existing motor pattern (Fig. 2c), or instruct motor learning. Importantly, these findings establish a movement signal that is probably compromised in Parkinson disease, and suggests the potential therapeutic use of restoring precisely timed, movement-associated dorsal striatum dopamine signalling[43].

Finally, distinct populations of axons signalling to reward or locomotion were observed both locally within dorsal striatum and globally across the striatum's dorsal–ventral axis suggesting regional tuning to regulate the specific functions of the terminal regions[34,39]. Heterogeneity of midbrain dopamine neurons has been described in terms of inputs[44,45], gene expression[46–48] and electrophysiological properties[47,49]; such cellular heterogeneity may align with the functional subpopulations described here (Supplementary Discussion)[39,44,45,50]. Specific targeting of these populations, and the local mechanisms such as cholinergic modulation[23,24] that shape their signalling, may provide novel therapeutic opportunities for neurological disorders.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 March 2015; accepted 20 June 2016.
Published online 11 July 2016.

1. Dauer, W. & Przedborski, S. Parkinson's disease: mechanisms and models. *Neuron* **39,** 889–909 (2003).
2. Hornykiewicz, O. [Dopamine (3-hydroxytyramine) in the central nervous system and its relation to the Parkinson syndrome in man]. *Dtsch. Med. Wochenschr.* **87,** 1807–1810 (1962).
3. Carlsson, A. Evidence for a role of dopamine in extrapyramidal functions. *Acta Neuroveg. (Wien)* **26,** 484–493 (1964).
4. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* **10,** 1020–1028 (2007).
5. Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips, P. E. & Graybiel, A. M. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* **500,** 575–579 (2013).
6. Romo, R. & Schultz, W. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* **63,** 592–606 (1990).
7. Pan, W. X., Schmidt, R., Wickens, J. R. & Hyland, B. I. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J. Neurosci.* **25,** 6235–6242 (2005).
8. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482,** 85–88 (2012).
9. Mirenowicz, J. & Schultz, W. Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* **72,** 1024–1027 (1994).
10. Hamid, A. A. *et al.* Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19,** 117–126 (2016).
11. Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology (Berl.)* **191,** 391–431 (2007).
12. Schultz, W. Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* **30,** 259–288 (2007).
13. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
14. Mazzoni, P., Hristova, A. & Krakauer, J. W. Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. *J. Neurosci.* **27,** 7105–7116 (2007).
15. Niv, Y., Daw, N. D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191,** 507–520 (2007).
16. Jin, X. & Costa, R. M. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* **466,** 457–462 (2010).
17. Puryear, C. B., Kim, M. J. & Mizumori, S. J. Conjunctive encoding of movement and reward by ventral tegmental area neurons in the freely navigating rodent. *Behav. Neurosci.* **124,** 234–247 (2010).
18. Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307,** 1642–1645 (2005).
19. Roesch, M. R., Calu, D. J. & Schoenbaum, G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.* **10,** 1615–1624 (2007).
20. Barter, J. W. *et al.* Beyond reward prediction errors: the role of dopamine in movement kinematics. *Front. Integr. Neurosci*. **9,** 39 (2015).
21. Dodson, P. D. *et al.* Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proc. Natl Acad. Sci. USA*, **113,** E2180–E2188 (2016).
22. Margolis, E. B., Lock, H., Hjelmstad, G. O. & Fields, H. L. The ventral tegmental area revisited: is there an electrophysiological marker for dopaminergic neurons? *J. Physiol. (Lond.)* **577,** 907–924 (2006).
23. Wang, L. *et al.* Modulation of dopamine release in the striatum by physiologically relevant levels of nicotine. *Nat. Commun.* **5,** 3925 (2014).
24. Threlfell, S. *et al.* Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron* **75,** 58–64 (2012).
25. Robinson, D. L., Venton, B. J., Heien, M. L. & Wightman, R. M. Detecting subsecond dopamine release with fast-scan cyclic voltammetry *in vivo*. *Clin. Chem.* **49,** 1763–1773 (2003).
26. Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. & Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nat. Neurosci.* **13,** 1433–1440 (2010).
27. Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499,** 295–300 (2013).
28. Matsuda, W. *et al.* Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *J. Neurosci.* **29,** 444–453 (2009).
29. Petreanu, L. *et al.* Activity in motor-sensory projections reveals distributed coding in somatosensation. *Nature* **489,** 299–303 (2012).
30. Cox, C. L., Denk, W., Tank, D. W. & Svoboda, K. Action potentials reliably invade axonal arbors of rat neocortical neurons. *Proc. Natl Acad. Sci. USA* **97,** 9724–9728 (2000).
31. Turner, T. J. Nicotine enhancement of dopamine release by a calcium-dependent increase in the size of the readily releasable pool of synaptic vesicles. *J. Neurosci.* **24,** 11328–11336 (2004).
32. Woodward, J. J., Chandler, L. J. & Leslie, S. W. Calcium-dependent and -independent release of endogenous dopamine from rat striatal synaptosomes. *Brain Res.* **473,** 91–98 (1988).
33. Tritsch, N. X., Ding, J. B. & Sabatini, B. L. Dopaminergic neurons inhibit striatal output through non-canonical release of GABA. *Nature* **490,** 262–266 (2012).
34. Parker, N. F. *et al.* Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* **19,** 845–854 (2016).
35. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8,** 1263–1268 (2005).
36. Ikemoto, S. Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res. Rev.* **56,** 27–78 (2007).
37. Hollerman, J. R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1,** 304–309 (1998).
38. Schultz, W. & Romo, R. Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *J. Neurophysiol.* **63,** 607–624 (1990).
39. Lerner, T. N. *et al.* Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. *Cell* **162,** 635–647 (2015).
40. Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76,** 470–485 (2012).
41. Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56,** 43–57 (2007).
42. Steinberg, E. E. & Janak, P. H. Establishing causality for dopamine in neural function and behavior with optogenetics. *Brain Res.* **1511,** 46–64 (2013).
43. Steinbeck, J. A. *et al.* Optogenetics enables functional analysis of human embryonic stem cell-derived grafts in a Parkinson's disease model. *Nat. Biotechnol.* **33,** 204–209 (2015).
44. Lammel, S. *et al.* Input-specific control of reward and aversion in the ventral tegmental area. *Nature* **491,** 212–217 (2012).
45. Beier, K. T. *et al.* Circuit architecture of VTA dopamine neurons revealed by systematic input-output mapping. *Cell* **162,** 622–634 (2015).
46. Poulin, J. F. *et al.* Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. *Cell Rep.* **9,** 930–943 (2014).
47. Roeper, J. Dissecting the diversity of midbrain dopamine neurons. *Trends Neurosci.* **36,** 336–342 (2013).
48. Lammel, S. *et al.* Diversity of transgenic mouse models for selective targeting of midbrain dopamine neurons. *Neuron* **85,** 429–438 (2015).
49. Matsumoto, M. & Hikosaka, O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459,** 837–841 (2009).
50. Gunaydin, L. A. *et al.* Natural neural projection dynamics underlying social behavior. *Cell* **157,** 1535–1551 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Graybiel and members of the Dombeck laboratory for comments on the manuscript, and V. Jayaraman, R. Kerr, D. Kim, L. Looger and K. Svoboda from the GENIE Project for GCaMP6. This work was supported by Klingenstein Foundation, McKnight Foundation, Whitehall Foundation, Chicago Biomedical Consortium with support from the Searle Funds at Chicago Community Trust, Northwestern University, National Institutes of Health (T32 AG20506).

Author Contributions M.W.H. performed the experiments, D.A.D. built the experimental apparatus, M.W.H. performed data analysis with strategy suggestions from D.A.D. Both authors conceived and designed the experiments, interpreted the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W.H. (markhowe72@gmail.com) or D.A.D. (d-dombeck@northwestern.edu).

METHODS

Virus injections for optical recording. All experiments were approved by the Northwestern University Animal Care and Use Committee. Heterozygous adult male mice (postnatal 3–4 months) with Cre expression in DAT-containing midbrain dopamine neurons (GENSAT line Tg(Slc6a3-cre)SG62Gsat/Mmucd) were anaesthetized with isofluorane (1–2%). For achieving widespread expression of GCaMP6f in dopaminergic neurons ($n=6$, Figs 1, 2, 4a, 5, Extended Data Figs 1, 2k, 3e, 4, 5, 10), a 0.5–1-mm diameter craniotomy was made over the right midbrain (−3.4 mm caudal, +1.0 lateral from bregma). A small volume (0.04 µl) of flexed-GCaMP6f virus (AAV1-Syn-flex-GCaMP6f, 1.4×10^{13} genome copies (GC) ml^{-1} diluted 1:1 in PBS; University of Pennsylvania vector core) was pressure injected through a pulled glass micropipette into the midbrain at two caudal locations (∼−3.3 and −3.5 mm from bregma) and 3 depths at each location (−4.1, −4.4 and −4.7 mm ventral from dura surface) for a total volume of ∼0.24 µl. For sparse, targeted injections (Fig. 4b–o, Extended Data Figs 2, 3, 7–9), the craniotomy was centred over either the medial VTA ($n=5$; caudal −3.1 mm, lateral 0.2 mm) or lateral SNc ($n=5$; caudal −3.4 mm, lateral 1.5 mm) and virus (0.025–0.05 µl) was injected at two ventral locations (−4.3, −4.6mm) for a total of ∼0.08 µl. Following the injections, the skull and craniotomy were sealed with Metabond (Parkell) and a custom metal headplate.

Dorsal striatum window implant surgery. One week after injection, mice were again anaesthetized, and the headplate and metabond removed. A ∼2.75 mm craniotomy was performed above the dorsal striatum (+0.5 mm rostral, +1.8–2.1 mm lateral to bregma) and cortical tissue was aspirated until the internal capsule fibres overlying the striatum were exposed[26]. Some of these fibres were then carefully removed until only a thin layer remained over the striatal surface. A thin (∼100 µm) layer of Kwik-Sil (WPI) was applied over the imaging region, and a metal cannula (stainless-steel tube, 2.7 mm outer diameter, 2.3 mm inner diameter, 1.6 mm length) covered at one end by a glass coverslip[26] was inserted into the aspiration site down to the fibre surface (Extended Data Fig. 2i). The imaging cannula, along with a metal headplate, and positioning ring (see ref. 26 for details) were sealed and attached to the skull with Metabond.

Behaviour and two-photon imaging. After the window surgery, the mice were allowed to recover in their home cages for ∼2–3 days. After this time, the mice began water scheduling (receiving ∼0.8–1 ml of water per day) as described previously[26,51]. The mice displayed no obvious signs of motor deficits following window implantation or water scheduling. Then 3–4 weeks after implant surgery, mice were head-fixed with their limbs resting on a 1D cylindrical styrofoam treadmill ∼20 cm in diameter by 13 cm wide (treadmill described previously[52], Fig. 1b) in the dark, which allowed them to run freely forwards and backwards. Large (16 µl) and small volume (4 µl) water rewards were delivered at pseudo-random time intervals (28.1 ± 19.8 s; mean ± s.d.) through a water spout gated electronically through a solenoid valve once imaging began, and spout licking was measured by a contact monitoring circuit. Solenoid valve triggering was accompanied by a short 'click' noise, which allowed us to conduct reward omission trials in which the water tube was manually closed (Extended Data Fig. 9a, b). Rotational velocity of the treadmill during locomotion was sampled at 1,000 Hz by a rotary encoder (E2-5000, US Digital used for most experiments; a few early experiments used a E6B2-CWZ3E, Yumo) attached to the axel of the treadmill. Instantaneous acceleration was calculated as the difference between consecutive treadmill velocity (first smoothed over two bins) measurements. Live video of behaviour on the treadmill was recorded at 30 Hz via a table mounted infared CCD camera (Computa) illuminated by an infared light source.

Two-photon imaging was performed using a custom, table mounted microscope system and data acquisition system described in ref. 53 (except the electric lens was not used here). Laser (920 nm) average power was in the range of 100–200 mW, but typically closer to 100 mW: This power measurement was made after the objective (Olympus LUMPlanFL N, 40×, 0.8 NA), but before the cannula, which apertures the excitation light, likely further reducing the power reaching the striatum. A Digidata1440A (Molecular Devices) data acquisition system was used to record (Clampex 10.3) and synchronize reward timing, licking, wheel velocity, and two-photon image frame timing. Imaging sessions began after mice were acclimated to head fixation and ran frequently on the treadmill (∼1–3 days). To rule out the influence of prior reward experience on locomotion responses, a subset of mice ($n=4$) were imaged before receiving any water rewards on the wheel (Extended Data Fig. 4h, i). Time-series videos (3,000–12,000 frames) were acquired at 28 Hz (256 × 64 pixels, 0.5 ms per line). A subset of the sparsely labelled single axon recordings were performed at 14 Hz (256 × 128 pixels, 0.5 ms per line) to achieve better spatial resolution for resolving small axonal arbors. For triggered averages (Fig. 4j, k, m), $\Delta F/F$ fluorescence traces generated from these 14 Hz imaging data sets were linearly interpolated to match the data collected at the 28 Hz sampling rates. Field of view sizes ranged from ∼60 to 250 µm.

Striatum imaging fields (60–250 µm diameter fields) were selected based on the presence of fibre morphology with at least occasional calcium transients in the fibres (field were not selected based on the behavioural correlation of the transients, only whether transients were clearly present) and, for widespread dopamine neuron expression experiments (Figs 1, 2, Extended Data Figs 1, 3, 4, 5), densely labelled projections across the field, or for sparse dopamine neuron expression experiments (Fig. 4, Extended Data Figs 7–9), few labelled projections in the field. Fields ranged in depth from 20–100 µm below the fibre surface. Each mouse was imaged for ∼5–10 days, with multiple fields imaged per day (an 'imaging session' consisted of time-series recording from a single field, typically for many minutes); ∼5–15 total different fields were imaged per mouse across all days (not all sessions met behaviour criteria to be included in all analyses, see below); no attempt was made to locate the same imaging field/axons from day to day, except in a small subset of experiments to demonstrate that axon morphology and behavioural signalling remains stable over days (Extended Data Fig. 2). Imaging up to ∼700 µm from the centre of the cannula was possible, although most fields were acquired near the centre of the cannula. Thus, most of our imaging fields were located more laterally in the dorsal striatum (∼1.5–2.4 mm lateral from the midline), but some medial fields were obtained as well (most medial, ∼1.1 mm from midline, Extended Data Fig. 1). Imaging along the rostral–caudal axis ranged from −0.2 to 1.2 mm from bregma. There was no preference for recording location within this large range of positions and no obvious differences in dopamine signalling were observed within this range.

Interpretation of mechanisms underlying calcium transients and characterization of putative single dopamine axon calcium transients. The time course of fluorescence changes caused by changes in cytoplasmic calcium concentration are well-known and characterized[27,30], providing a significant framework within which to interpret our observed calcium transients (Extended Data Fig. 3a–c). For example, a short tens of milliseconds timescale influx of calcium (such as a short burst of action potentials over tens of milliseconds) results in a stereotyped calcium transient consisting of a fluorescence increase to peak over ∼50–70 ms followed by an exponential decay to baseline over ∼150–250ms (Extended Data Fig. 3a). Multiple influx events separated by less than the indicator decay time (for example, longer bursts of action potential firing over ∼100s of milliseconds) lead to longer duration, larger amplitude calcium transients due to summation (Extended Data Fig. 3b), while a sustained increase in the rate of influx events separated by less than the decay time (for example, sustained increase in action potential firing) leads to sustained elevations or apparent shifts in the fluorescence baseline due to ongoing summation[54] (Extended Data Fig. 3c, left). Thus, GCaMP6f fluorescence changes are well-suited for determining whether mouse locomotion (accelerations on the 100s of millisecond timescale) is associated with sub-second phasic dopamine signalling (rapid transient changes in fluorescence), slowly varying dopamine signalling (long-duration transients or changes in the GCaMP6f baseline level over many seconds or minutes) or both. However, our imaging methods probably cannot reliably detect single action potentials in single axons, or the difference in transient amplitude due to an additional spike in a burst. Furthermore, it is possible that small, slow changes in action potential firing could occur between rest and locomotion and not be detected using our optical recording methods. Also, although we did not detect slow changes in signalling here, such changes could occur over timescales longer than our typical recording sessions (many minutes).

Optical stimulation of dopamine axons. To achieve widespread expression of ChR2 in dopaminergic axons in the striatum, DAT-cre mice ($n=7$; 5 male, 2 female; postnatal 3–6 months) were injected bilaterally (0.2–0.3 µl total per hemisphere) as above with AAV9-EF1a-DIO-hChR2 (University of Pennsylvania vector core diluted 3:1 in PBS to 3.5×10^{12} GC ml^{-1}) in the midbrain (−3.3 mm caudal, ±1.3 mm lateral from bregma) at three depths (−4.1, −4.4, −4.7 mm from the dura surface). Headplates for head fixation were then attached to the skull with a thin layer of Metabond and the location of the striatum (+0.5 mm rostral, ±1.8 mm lateral from bregma) was marked on the surface. After 4–6 weeks[55], mice were anaesthetized with isoflourane and small craniotomies (∼0.5–1 mm in diameter) were drilled (at the pre-marked locations) through the Metabond and skull over the striatum, leaving the dura and cortex intact. The craniotomy was then sealed with Kwik-Sil. After the mice recovered from this short (∼10–15 min) surgery for at least one day, they were head-fixed on the linear treadmill. When mice began to exhibit periodic, spontaneous run-rest transitions on the treadmill, the Kwik-Sil covering the craniotomies was removed and two optical fibres (200 µm diameter, ThorLabs) were lowered slowly into the brain bilaterally (to various depths ranging from dorsal (1.6–2.2 mm) to ventral (3.6–4.1 mm) striatum) to allow for delivery of laser light (488 nm diode laser, Thor Labs). Laser stimulation (5 mW laser power at the fibre tip) trains consisted of 83 ms ON/83 ms OFF (6 Hz) or 166 ms ON/166 ms OFF (3 Hz) periods and the laser was pulsed (8 ms ON/8 ms OFF, using an optical chopper, ThorLabs) during the 83 or 166 ms ON periods.

We refer to this stimulation as a train of 8-ms pulse bursts repeated at 3 or 6 Hz. For example, for 6-Hz stimulation, the laser stimulation trains consisted of the following: 8 ms ON/8 ms OFF/8 ms ON/8 ms OFF (repeated for 83 ms) followed by 83 ms with no stimulation and then the whole sequence repeated (Fig. 3b). The onset of laser stimulation trains were triggered manually via a pulse generator and synchronized to output from the treadmill rotary encoder through our data acquisition system (see above). Stimulation trains (duration 8.6 ± 2.8 s) were initiated periodically at pseudo-random intervals primarily during rest periods during 1–8 stimulation blocks (sessions) per mouse per day at varying depths in the striatum (3–5 min block durations, all occurring during a single ~40 min recording period on the treadmill per day per mouse). Importantly, bright LED lights (470 nM and 530 nM ThorLabs) provided constant illumination onto the face of the mice on the treadmill during the stimulation blocks to mask the mice from the fibre optic light delivered to their brain. After stimulation was completed, craniotomies were again covered with Kwik-Sil to allow for additional sessions on subsequent days (1–3 days per mouse). We note that effects highly similar to those shown in Fig. 3 and Extended Data Fig. 6a–d were obtained using the above laser stimulation trains, but with either faster pulsing (4 ms ON/4 ms OFF) during the 83 or 166 ms ON periods or no pulsing at all (only 83 ms ON/83 ms OFF, 6 Hz or 166 ms ON/166 ms OFF, 3 Hz); however all data presented in the manuscript were obtained using the 8-ms pulsing during bursts protocol.

Single fibre photometry. For fibre photometry, we coupled the blue diode laser excitation light (488 nm diode laser, Thor Labs; filtered using a 488/6 nm filter, Semrock) into an optical fibre (200 μm diameter, ThorLabs) and used neutral density filters between the laser and fibre coupling lens to adjust power (1 mW at fibre tip). Fluorescence was collected through the same fibre and was separated from the excitation light by placing a dichroic mirror (505 nm-shortpass, Semrock) between the laser and fibre coupling lens. Fluorescence light was further filtered (540/50 nm, Semrock) and then measured using a GaAsP PMT (H10770PA-40, Hamamatsu). The PMT signal was digitized and recorded using the electronics and ScanImage software from our two-photon microscope (described above). Synchronization of the recorded fluorescence with treadmill movements was accomplished as described above for two-photon imaging.

DAT-cre mice ($n = 5$) were injected with viral vectors to obtain ubiquitous GCaMP6f expression in midbrain dopamine neurons as described above for two-photon imaging experiments and fitted with plates for head-fixation as described for the ChR2 experiments. At 3–5 weeks after injection, mice were trained for 1–3 days on the linear treadmill with rewards until they ran reliably. Mice were then anaesthetized with isoflourane and a unilateral craniotomy (~0.5–1 mm in diameter, dura and cortex were left intact) over the central striatum (+0.5A/P, +1.3–1.8mm M/L from bregma) was performed. After recovery from the short (~10–15 min) surgery, mice were head-fixed on the linear treadmill. When mice began to run regularly on the treadmill, the optical fibre was lowered slowly into the brain to various depths ranging from Cortex to Ventral Striatum for photometry recording (Fig. 5, Extended Data Fig. 10). Unpredicted rewards were delivered as described above.

Data analysis. Data were analysed using ImageJ (1.46) and custom functions written in MATLAB (Version 2012b). Sample sizes were chosen to reliably measure experimental parameters while remaining in compliance with ethical guidelines to minimize the number of animals used. Experiments did not involve blinding, but randomization was used with respect to reward delivery and order of data collection from control versus experimental mice. All data in the text and figures are labelled as either mean ± s.d. or mean ± s.e.m.

Motion correction and ROI selection for two-photon microscopy. Each time-series video was motion corrected using algorithms described previously[26,56]. Videos were then visibly inspected to confirm post-correction image stability and videos with excessive motion artefact or apparent z-shifts were excluded from analysis. To quantify whole-field fluorescence (Figs 1, 2, Extended Data Figs 1, 2k, 3e, 4, 5) a single large region of interest (ROI) was hand-selected (to contain all the active regions of the field) from the mean fluorescence image for each time series. Fluorescence traces and ROIs from putative single axon segments from the sparsely labelled fields (Fig. 4, Extended Data Figs 7–9) were generated using CellSort[57] with the following parameters: smoothing width = 1.5, mu = 0.8, principal components = 200, area limits = 50–3,000 pixels. ROI fluorescence traces and ROI morphology were visibly inspected and selected for inclusion as a putative axon if they exhibited elongated morphology consistent with axons (Fig. 4c) and activity patterns with predominantly positive-going fluorescence changes (no structures with slowly varying fluorescence activity patterns were observed). The fluorescence traces were then baseline normalized (eighth percentile fluorescence over a 1,000 bin, ~35.7 s, sliding window, to remove any slow drifts in baseline; see below for measurements of slow changes in baseline) and converted to $\Delta F/F$. Significant positive-going transients (used for Figs 1g, 2g, 4k–o; Extended Data Figs 2g, h, 3d, e, 4h, i (bottom), 5a, b, 7–9, 10a–c) were identified as previously described[26]. To account for axon segments that probably originated from the same neuron, hierarchical clustering (simple linkage) of the linear correlation matrix between $\Delta F/F$ traces of all segments in each field was performed using a correlation coefficient threshold of 0.8 (results were similar using other clustering thresholds, Extended Data Fig. 8a, b), and $\Delta F/F$ traces belonging to the same cluster were averaged. To verify single axon separation from background, analysis was repeated only on the most sparsely expressing fields; separation of functional axon classes was qualitatively similar to Fig. 4n using this more conservative approach (data not shown).

Relationship of fluorescence signals to treadmill locomotion. All mice ($n = 6$ whole field, $n = 5$ for each single-axon imaging group, $n = 5$ for photometry) contributed to averaged measures unless otherwise noted in the text or figure captions. Individual imaging sessions were excluded if they did not meet criteria for each specific behavioural analysis (that is, locomotion initiation, continuous locomotion, locomotion terminations; see below for each), hence the session numbers varied slightly for each measure. For general comparisons of fluorescence changes in whole fields and single axons during locomotion and rest periods (Figs 1f, g, 4n, o, Extended Data Figs 2f, 4h, 7, 8,10a–c, g), we classified each imaging frame (~35-ms bin size) as occurring during active locomotion, passive rest, or neither (uncertain, excluded). All bins occurring within 5.5 s post-reward delivery or during a spontaneous licking period were excluded. Locomotion bins were identified as those that had a treadmill velocity measure greater than $0.2\,\mathrm{m\,s^{-1}}$ and an acceleration greater than $1.7\,\mathrm{m\,s^{-2}}$ in a 200-ms window on either side. Conversely, rest bins were defined as having no accelerations greater than $0.3\,\mathrm{m\,s^{-2}}$ or velocities greater than $0.05\,\mathrm{m\,s^{-1}}$. These parameters were determined based upon visual comparison of live video with acceleration and velocity traces and are probably conservative constraints. To be included for locomotion/rest comparisons, behaviour over the entire imaging session had to include a minimum of 30 rest bins and 30 locomotion bins. Slow, sustained changes in baseline fluorescence (Figs 1f and Extended Data Fig. 10a–c) were quantified by comparing $\Delta F/F$ for locomotion and resting periods that did not contain significant calcium transients. The $\Delta F/F$ baseline normalization was re-computed for this calculation using a larger sliding window (2,000 bins, ~70.4 s) to ensure that slow events over multiple locomotion bouts (Extended Data Fig. 3e) would be detected if they were present. Further, examination of $\Delta F/F$ traces with no baseline normalization did not reveal slow events corresponding to locomotion bouts (Extended Data Fig. 2k).

Locomotion initiation times for triggered averages of fluorescence (Figs 2a, d, f, 4h, Extended Data Figs 4a, b, j, 10d–f) were first identified manually by visual inspection of velocity traces for each session. Typically, mice exhibited clear transitions between rest (periods with few treadmill accelerations) and continuous treadmill locomotion. Bursts of acceleration not followed by a prolonged elevation in velocity (indicative of continuous locomotion) were not selected. Manually selected initiations were included for analyses only if they were followed (within 10 bins, ~350 ms) by an acceleration peak greater than $1\,\mathrm{m\,s^{-2}}$ and were not preceded within 1 s by any velocity bins larger than $0.05\,\mathrm{m\,s^{-1}}$ or accelerations larger than $0.3\,\mathrm{m\,s^{-2}}$. Acceleration-triggered averages were calculated relative to the onset of the first acceleration (after the manually defined time point) greater than $0.3\,\mathrm{m\,s^{-2}}$. Locomotion terminations (Fig. 2c) were also selected manually as the first deceleration to rest following locomotion bouts. One termination time-point was selected for each initiation. Terminations were included if they were preceded (within 10 bins) by a maximum velocity greater than $0.1\,\mathrm{m\,s^{-1}}$, indicating a sharp locomotion to rest transition. For defining periods of acceleration occurring during continuous running (Fig. 2b, e, f, Extended Data Figs 4c, d, k, m, 8d, e), accelerations were required to be preceded (within a ~200-ms window) by a minimum run velocity of ($0.1\,\mathrm{m\,s^{-1}}$) and followed by a mean run velocity (over 200 ms) of $0.1\,\mathrm{m\,s^{-1}}$. Non behaviour-specific acceleration triggered averages of fluorescence (Figs 4j, m, 5d, Extended Data Fig. 4i, top) were calculated for all accelerations exceeding $1.7\,\mathrm{m\,s^{-2}}$ (that is, accelerations occurring during continuous locomotion, locomotion onsets and any other short movements not falling into either of these two categories, though the vast majority occurred during continuous locomotion).

For acceleration triggered averages of $\Delta F/F$, single average $\Delta F/F$ traces were derived for each field (Figs 1 and 2), putative single axon (Fig. 4), or photometry recording location (Fig. 5, Extended Data Fig. 10) across all identified accelerations (meeting acceleration criteria, triggered on the zero crossing of the acceleration trace when the acceleration was increasing and crossed a threshold value $>1.7\,\mathrm{m\,s^{-2}}$) in an imaging session; and these traces were averaged across sessions and mice to generate summary traces. $\Delta F/F$ triggered averages of acceleration were calculated analogously (Fig. 2g, Extended Data Fig. 4i, bottom, triggered on the onsets of significant, positive-going $\Delta F/F$ transients). For these traces, and for analysis of amplitudes of preceding and following accelerations bursts (Extended Data Fig. 5c, d) only significant transients <0.5 s duration were included to avoid

inclusion of any sustained summated transients. Shuffled triggered averages were computed relative to an equivalent number of randomly selected time-points (randsample in MATLAB). Correlations (Pearson's) between acceleration and $\Delta F/F$ were calculated between the (concatenated) traces around all initiation bins ($\pm$750 ms around each) or continuous locomotion periods (+2 s after locomotion initiation to locomotion termination). Note that differences in the criteria for selecting different behaviour periods (locomotion initiations, terminations, continuous running periods, etc) results in the inclusion of a different number of fields or axons for each behaviour (that is, Fig. 2a–e).

Single-photon fibre photometry fluorescence data was first averaged over 35.7-ms bins and then corrected for background signal (intrinsic fluorescence and laser bleed-through). The signal recorded in the unlabelled cortex (estimated in a subset of the recordings) was assumed to be background and this signal was subtracted from the striatum signal. After this subtraction, the remaining photometry signal (assumed to be GCaMP6f fluorescence) was analysed identically to the whole field two-photon imaging data sets described above.

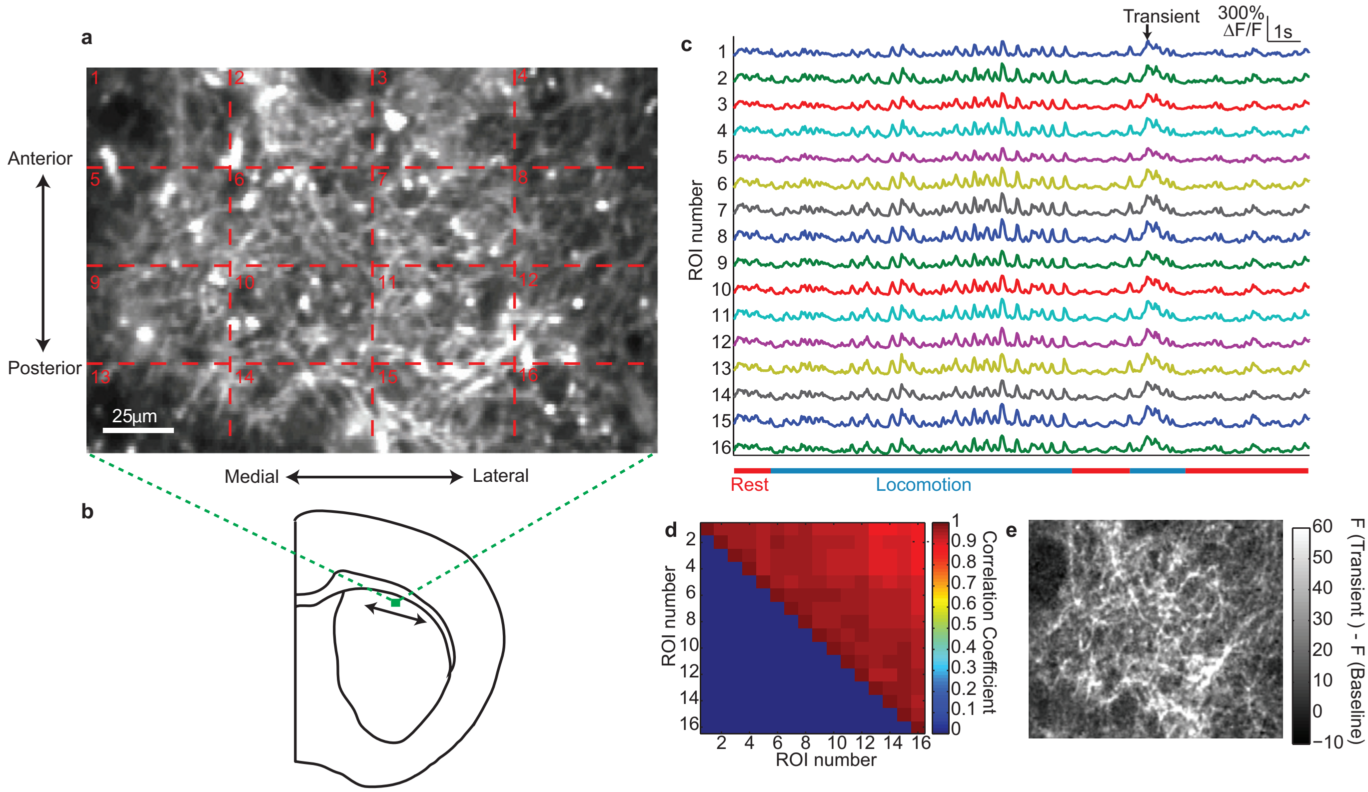
Relationship of acceleration to ChR2 axon stimulation. Significant effects of laser stimulation on locomotion initiations (Fig. 3d–g) were assessed by comparing the mean absolute value of acceleration within a 0.5–2-s window (for each mouse and across all mice, Fig. 3f) or for each bin (Fig. 3g) following stimulation train onsets that were delivered during rest (no large accelerations within 2 s before stimulation) across all mice and stimulation trains to the means obtained from 1,000 rounds of randomly selecting an equivalent number of non-laser stimulation rest periods from the same stimulation sessions. Sessions were included if mice exhibited spontaneous locomotion/rest transitions in the absence of stimulation and exhibited average resting bout durations of >6 s (to allow comparison with spontaneous transitions from rest to running). Stimulation effects were considered significant if the post-stimulation means were greater than 990 of the 1,000 randomly selected means ($P < 0.01$). Note that because we analysed laser train onsets that were preceded by 2 s of rest, there was a slightly higher probability for mice to be moving in a window after the stimulation onset simply by chance. This explains the apparent small change in the control group (Fig. 3f) at laser train onset and in the shuffle group (Fig. 3g). A similar test was applied to assess significant frequency dependent acceleration entraiment (Extended Data Fig. 6c, d). Laser burst triggered averages of acceleration were computed during locomotion periods beginning after laser train onset and the mean of absolute value of these triggered averages were compared with 1,000 averages from randomly selected non-laser stimulation locomotion periods. Consistent phase relationships between the laser stimulation and acceleration produce significant positive and negative peaks in the triggered average (Extended Data Fig. 6c,d), whereas traces with no consistent phase relationship result in flat triggered averages (see controls, Extended Data Fig. 6c, d). Multi-taper power spectral densities of acceleration (Fig. 3h) were constructed for all locomotion periods (minimum 3 s of locomotion per stimulation train period) during 6 Hz and 3 Hz stimulation periods in each block using the mtspectrumc function in the Chronux library (http://chronux.org/) with 5 tapers and a time bandwidth product of 3 over the interval 1 Hz to 8 Hz. Power spectra across blocks were averaged to produce mean power spectra for each mouse (one each for 6 and 3 Hz stimulations, Fig. 3h). The centre of mass (Fig. 3i–l, Extended Data Fig. 6f–h) was calculated from the top 25% of the mean power spectra for each mouse (although highly similar results were obtained using the top 50%, 75% or all of the power spectra).

Classification of functional signalling related to reward and locomotion across putative single axons and striatum locations. Axon heterogeneity in single fields relative to reward and locomotion periods (Fig. 4a, i) was visualized by computing the mean fluorescence image during locomotion bins (see above) and subtracting the mean fluorescence image during reward delivery (a 750-ms window following reward delivery). Locomotion index (Fig. 4n and Extended Data Figs 7, 8, 10g) was the difference between the mean $\Delta F/F$ during locomotion and rest ($\Delta/\Delta F_{\text{move}} - \Delta/\Delta F_{\text{rest}}$) and axons were classified as significantly locomotion responsive via a Wilcoxon rank-sum test ($P < 0.05$) between all locomotion and rest $\Delta F/F$ values. Reward responses (Fig. 4n and Extended Data Figs 7, 8, 10g) were defined as the mean $\Delta F/F$ in a 300-ms window after reward delivery. Significance of reward responses were computed with a bootstrap test ($P < 0.05$) where the P value was the probability of receiving an equivalent or larger mean value from 1,000 rounds of randomly selecting a trigger bin from all non-reward bins.

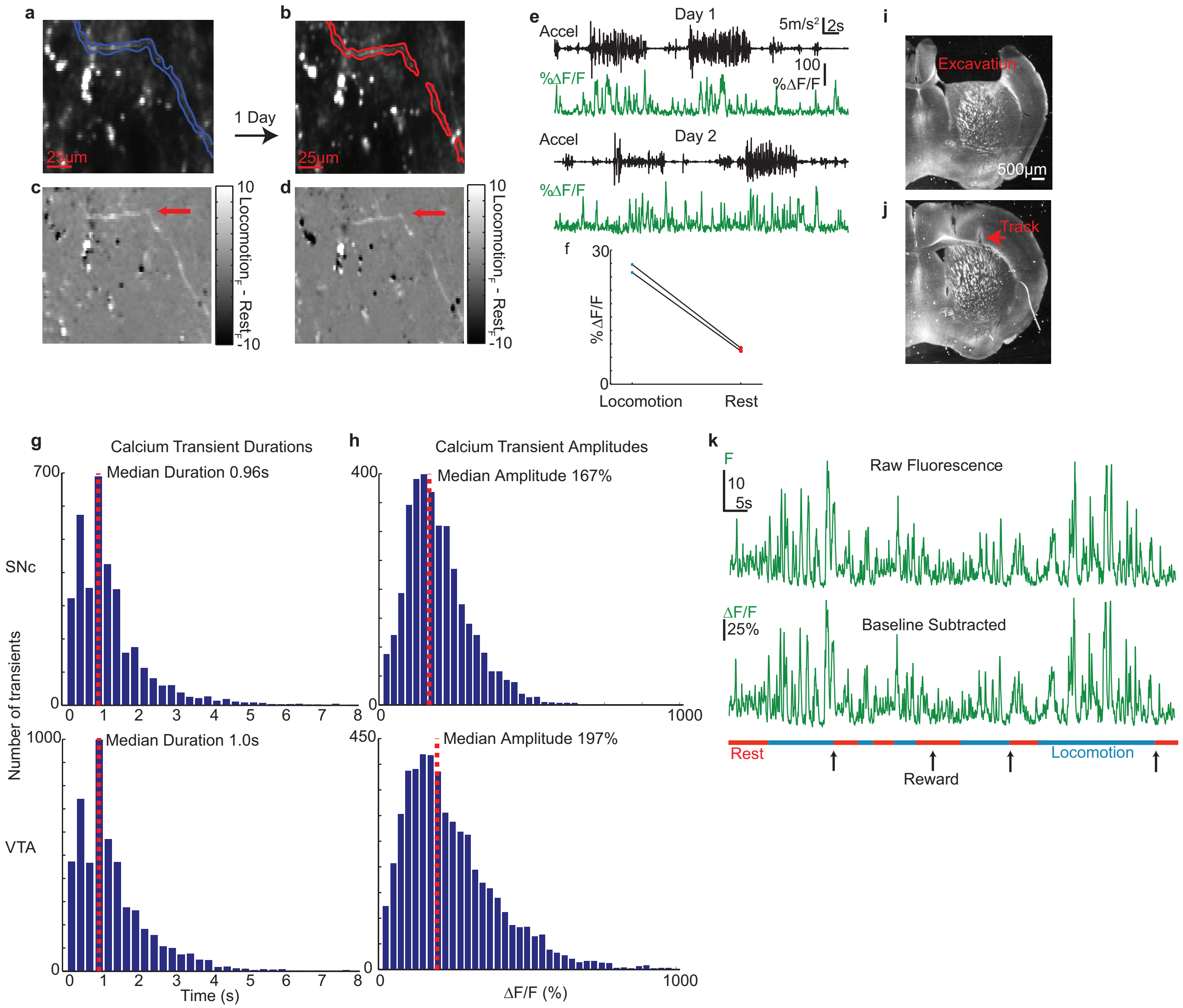
ROC analysis. ROC (receiver operating characteristic) curves (Extended Data Fig. 4n, o) to assess the ability of calcium transients to discriminate locomotion from rest were computed using the 'perfcurve' function in MATLAB for all bins for each imaging field that were classified as locomotion or rest based on the criteria defined above. The positive state was defined as locomotion, the negative as rest (Extended Data Fig. 4n). For the onset ROC analysis to assess ability of calcium transients to discriminate pre-locomotor from other rest periods (Extended Data Fig. 4o), all resting bins were classified as either being pre-onset (positive state, within 250 ms before defined onset; see above) or non-pre onset (negative state).

Histology. Mice with injections targeting SNc or VTA (see above) were perfused trans-cardially with 15ml PBS (Fischer) and 15 ml of 4% paraformaldehyde (EMS). Brains were stored in PBS at 4 °C then transferred to 40% sucrose (Fischer) overnight before sectioning. Coronal slices (40–50 μm) were cut on a freezing microtome and stored at 4 °C in PBS. For immunostaining of dopaminergic neurons, sections from a subset of mice were blocked in 5% serum, incubated overnight at 4 °C with antibodies for tyrosine hydroxylase (1:700 dilution, Millipore AB152, Fig. 1d, Extended Data Fig. 7), then incubated with secondary antibodies tagged with Alexa Fluor 594 (Jackson). Large area imaging of tyrosine hydroxylase and GCaMP6f expression was performed on an Olympus Slide Scanner (VS120) microscope, and high magnification imaging of co-labelling (Fig. 1d, right) was performed with a confocal microscope.

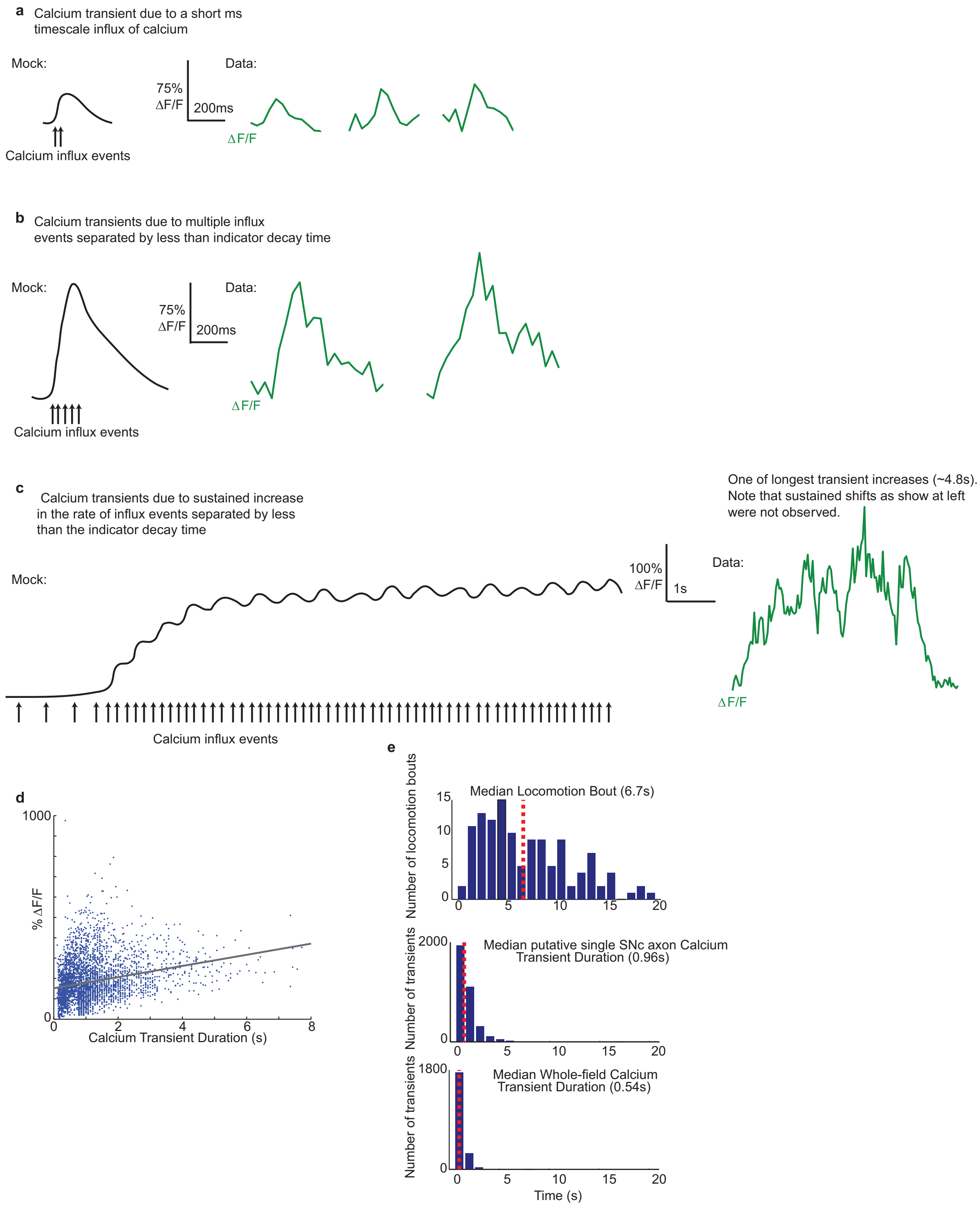
51. Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461,** 941–946 (2009).
52. Heys, J. G., Rangarajan, K. V. & Dombeck, D. A. The functional micro-organization of grid cells revealed by cellular-resolution imaging. *Neuron* **84,** 1079–1090 (2014).
53. Sheffield, M. E. & Dombeck, D. A. Calcium transient prevalence across the dendritic arbour predicts place field properties. *Nature* **517,** 200–204 (2015).
54. Hofer, S. B. *et al.* Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat. Neurosci.* **14,** 1045–1052 (2011).
55. Miyashita, T., Shao, Y. R., Chung, J., Pourzia, O. & Feldman, D. E. Long-term channelrhodopsin-2 (ChR2) expression can induce abnormal axonal morphology and targeting in cerebral cortex. *Front. Neural Circuits* **7,** 8 (2013).
56. Miri, A. *et al.* Spatial gradients and multidimensional dynamics in a neural integrator circuit. *Nat. Neurosci.* **14,** 1150–1159 (2011).
57. Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* **63,** 747–760 (2009).



Extended Data Figure 1 | Synchronized dopamine projection axon dynamics across a single field in dorsal striatum. a, Representative mean fluorescence image in dorsal striatum of a dense field of dopamine axons (compare to the sparse fields from sparse labelling in Fig. 4c, i) from one mouse (out of six) labelled with GCaMP6f. **b**, Coronal schematic showing approximate location and scale of region imaged at top. Arrow indicates the approximate range of medial/lateral positions used for two-photon imaging (see Methods). **c**, $\Delta F/F$ traces for the rectangular ROIs indicated in **a**. **d**, Correlation matrix for the ROIs indicated in **a**. Note high degree of transient co-activation across ROIs. **e**, Mean image of fluorescence during the transient indicated by the arrow in **c** minus mean image of fluorescence during non-transient periods for the field shown in **a**. Note that the morphology of active regions closely resembles the morphology of GCaMP6f-expressing axons in the whole field in **a**, indicating synchronous activation of large, dense regions of axons, likely belonging to several different parent neurons.

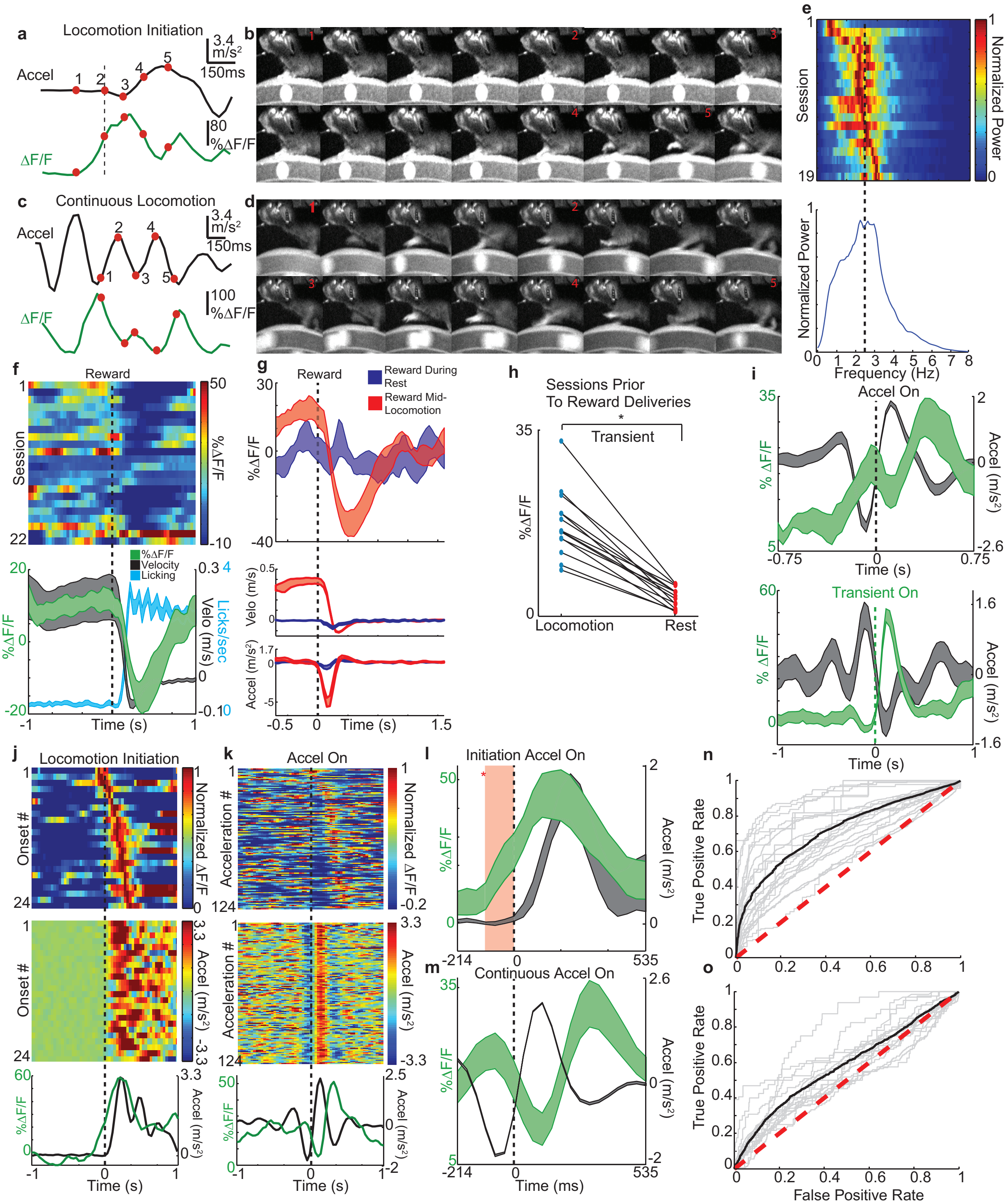


Extended Data Figure 2 | Further characterization of two-photon imaging and analysis methods. **a**, **b**, Top, example mean fluorescence images of a putative single SNc axon imaged over 2 consecutive days in one mouse. **c**, **d**, Mean images of fluorescence during locomotion periods minus mean image of fluorescence during reward periods for fields in **a** and **b**. White axonal regions indicate regions of elevated signalling during locomotion. Note the similar morphology and behaviour signalling of the identified axon (red arrow) over days. **e**, Acceleration (black) and $\Delta F/F$ (green) for the identified axon in **a**–**d** across the two imaging days. Note the similar transient amplitudes and the elevated transient signalling during locomotion acceleration periods. **f**, Mean transient $\Delta F/F$ (mean of significant transients, excluding baseline periods) during locomotion and rest on days 1 and 2 for the axon shown in **a**–**d**. **g**, Histograms of calcium transient duration times across all putative single axons imaged in dorsal striatum from SNc ($n = 3{,}556$ transients, 5 mice top) and VTA ($n = 5140$ transients, 5 mice bottom). Note the similar duration profile across the two populations (medians not significantly different, $P > 0.05$ Wilcoxon sign-rank test). **h**, Histograms of maximum calcium transient amplitudes across all putative single axons imaged in dorsal striatum from SNc (top) and VTA (bottom) (medians not significantly different, $P > 0.05$ Wilcoxon sign-rank test). **i**, Post-mortem image of a coronal section from a representative mouse showing the striatum imaging cannula window cortical lesion site. **j**, Post-mortem image from a different mouse than **i** that was used for fibre photometry recording (fibre track indicated by arrow). **k**, Similarity of raw (no baseline normalization) whole-field GCaMP6f F trace (top) with baseline normalized $\Delta F/F$ trace (bottom) for the example whole-field imaging session shown in Fig. 1e. In the top trace, note the lack of baseline change over the recording session and particularly the stability of the baseline level during locomotion periods. Bottom trace is duplicate of trace in Fig. 1e.



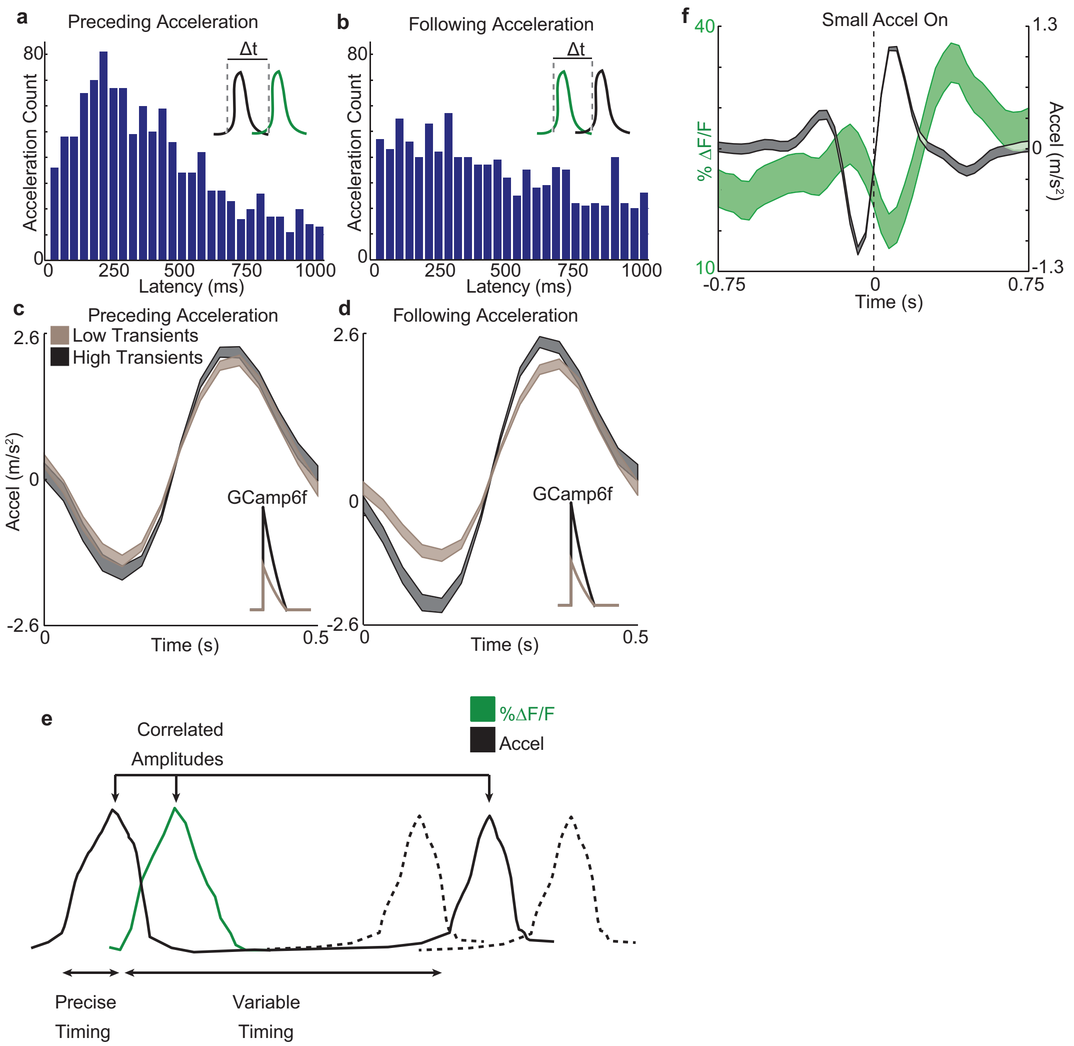
Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Interpretation of mechanisms underlying calcium transients and characterization of putative single dopamine axon calcium transients. a, Left, mock trace representing expected GCaMP6f calcium transient from a short millisecond timescale influx of calcium (arrows; for example, short burst of action potentials over tens of milliseconds; local modulation may also contribute to calcium influx). Right, three representative low-amplitude, short duration calcium transients (from putative single dopamine axons in dorsal striatum) (see Methods, Fig. 4) that display onset and decay kinetics consistent with mock transient (left). **b**, Left, mock trace representing expected GCaMP6f calcium transient from multiple calcium influx events separated by less than the indicator decay time (arrows; for example, longer burst of action potential firing over ~100s of milliseconds). Right, representative larger amplitude calcium transients (from putative single dopamine axons in dorsal striatum) with rapid rise times consistent with mock transient (left). **c**, Left, mock trace representing expected GCaMP6f calcium transient from a sustained increase in the rate of influx events separated by less than the decay time (for example, sustained increase in action potential firing). Right, representative trace of one of the longest duration calcium transients observed (from putative single dopamine axon in dorsal striatum). Note that no sustained increases (baseline shifts) similar to the mock trace (left) were observed in single axon recordings or whole-field $\Delta F/F$ measurements; however, the long duration transient shown (right) indicates that if such sustained increases had occurred, they would have been detected using our methods. Also, note that the mock traces shown in **a**–**c** are for descriptive purposes and are not based on new data. These traces are based on two main assumptions: (1) $\Delta F/F$ is a monotonically increasing function of intracellular calcium concentration, which itself is a monotonically increasing function of the number of underlying action potentials (that is, a greater number of action potentials leads to a larger $\Delta F/F$, but the relationship is not necessarily linear); and (2) $\Delta F/F$ transients summate (not necessarily linearly) when they overlap in time. **d**, Duration versus peak $\Delta F/F$ for all identified significant calcium transients in putative single SNc originating axons (see Methods, Fig. 4, $n = 3{,}556$ transients from 73 axons in 5 mice; Spearman's Rho $= 0.3$ $P < 10^{-10}$). **e**, Histogram of sustained locomotion period durations (from SNc injected mice, $n = 5$, top) and calcium transient durations for all putative single SNc axons ($n = 5$ mice, mid) and all whole-fields ($n = 6$ mice, bottom). Note that the median calcium transient duration (for either single axon or whole-field) is far less than the median locomotion duration, indicating that the increase in dopamine axon GCaMP6f $\Delta F/F$ observed during locomotion is due to an increase in relatively short duration calcium transients, rather than long-duration (sustained) increases in $\Delta F/F$.

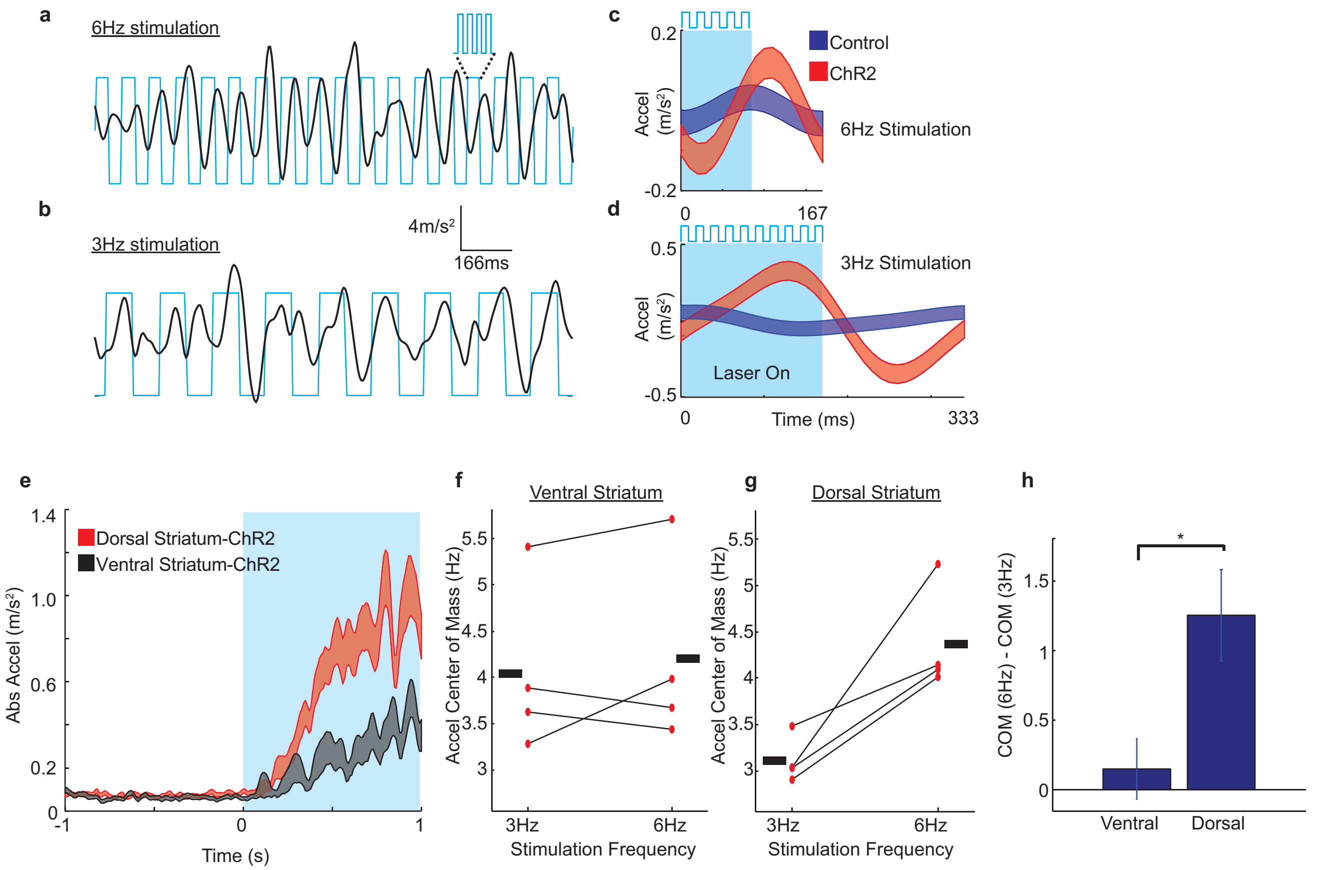


Extended Data Figure 4 | See next page for caption.

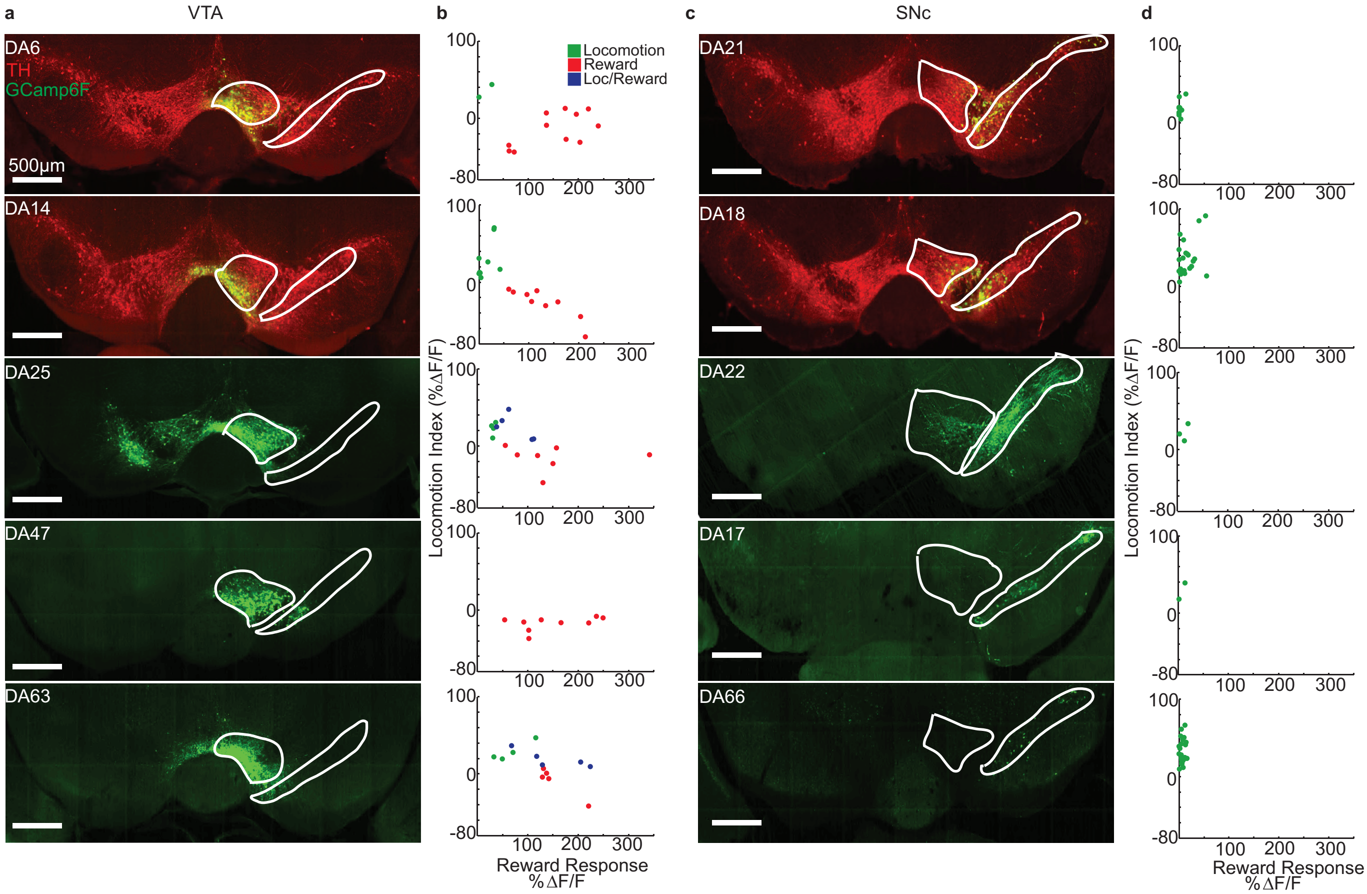
Extended Data Figure 4 | Further characterization of acceleration-associated dopamine signalling. a, Representative whole-field $\Delta F/F$ fluorescence trace (one field from one out of six mice, green) aligned to treadmill acceleration (black) during a locomotion onset (dashed line: onset). **b**, Video frames of mouse for time points shown in **a**. **c**, **d**, Same as **a** and **b** but for a period of continuous locomotion. **e**, Top, normalized spectral power of treadmill acceleration trace during continuous locomotion periods for each two-photon imaging session (each row represents a session, $n = 6$ mice). Bottom, normalized mean power from all sessions shown in **e**, top. **f**, Top, mean whole-field $\Delta F/F$ trace triggered on reward delivery time for all fields; each row is mean for each field/session ($n = 22$ fields, 6 mice). Bottom, mean treadmill velocity (black), mean whole-field $\Delta F/F$ (green), and spout licking (light blue) all triggered on reward time (mean across all 22 fields/sessions in 6 mice). **g**, Mean whole-field $\Delta F/F$ (top), velocity (mid), and acceleration (bottom) for trials in which reward was delivered mid-locomotion ($n = 13$ sessions, 4 mice, red) or when animals were at rest ($n = 12$ sessions, 4 mice, blue). Note the sharp decrease in $\Delta F/F$ relative to baseline when animals decelerated from locomotion to consume the reward and the relative absence of phasic reward signalling when animals were given reward from rest. Reward responses were also not observed in single SNc axons when animals received reward from rest (see Extended Data Fig. 9). **h**, Comparison of mean whole-field fluorescence change from significant calcium transients (excluding baseline periods) between locomotion and resting periods; each point represents mean $\Delta F/F$ for running or resting over one session for each field (lines connect same field/session). All fields included here were imaged before mice ever receiving any rewards on the treadmill ($n = 14$ fields, 4 mice). $*P < 10^{-5}$ (Wilcoxon rank-sum test). **i**, top, Mean acceleration (black) and whole-field $\Delta F/F$ (green) triggered on acceleration onsets (mean across all fields) during continuous locomotion. Bottom, mean acceleration (black) triggered on all short duration calcium transients (green, mean of transients) during continuous locomotion across all fields. All fields included here were imaged before mice ever receiving any rewards on the treadmill ($n = 14$ fields, 4 mice). **j**, Top, whole-field $\Delta F/F$ from all locomotion initiations in a representative single session (single imaging field, single session, one out of 6 mice); each row represents a single locomotion initiation time period (sorted by peak $\Delta F/F$ time). Mid, treadmill accelerations corresponding to locomotion initiations shown in **j**, top. Bottom, average of acceleration (black) and $\Delta F/F$ (green) across all locomotion onset traces displayed in Top and Mid panels. **k**, same as **j**, but for continuous locomotion periods. **l**, Reproduction of Fig. 2a with zoomed-in time axes to show the timing of the mean DF/F in relation to the first acceleration at locomotion initiations from rest. Shaded red region indicates bins that were significantly ($*P < 0.01$, Wilcoxon sign-rank, $n = 15$ fields in 6 mice) elevated relative to rest baseline. Shaded region covers ~107 ms before acceleration onset. **m**, Reproduction of Fig. 2b with zoomed-in time axes to show the timing of the mean dopamine transient in relation to the accelerations during continuous locomotion ($n = 18$ fields, 6 mice). **n**, ROC curves for each two-photon whole-field $\Delta F/F$ trace ($n = 22$ fields from 6 mice, grey; mean, black line) assessed for ability to discriminate locomotion versus resting periods (21 out of 22 exhibited significant discriminability, $P < 0.01$). Area under the curve (AUC) $= 0.76 \pm 0.02$ (mean ± s.e.m.). **o**, ROC curves for each two-photon whole-field $\Delta F/F$ trace ($n = 17$ fields in 6 mice, grey; mean, black line) assessed for ability to discriminate pre-locomotion onset rest periods (250 ms before onset) from other rest periods (10 out of 17 exhibited significant discriminability, $P < 0.01$, two sessions included did not meet onset criteria for Fig. 2a, Methods). AUC $= 0.58 \pm 0.02$ (mean ± s.e.m.). Dashed red lines indicate the line of no discrimination. Shaded regions in **f**, **g**, **i**, **l**, **m** denotes mean ± s.e.m.



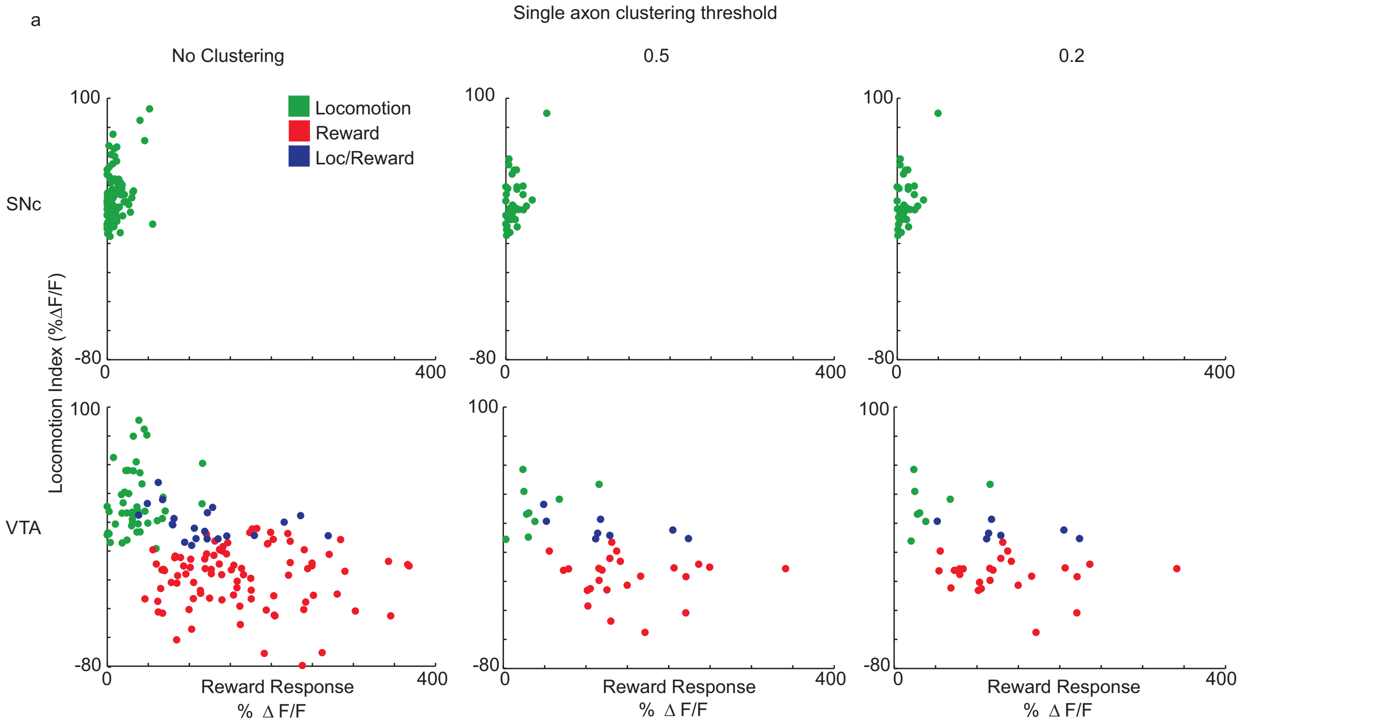
Extended Data Figure 5 | Dopamine axon calcium transients are temporally associated with preceding acceleration bursts and their amplitude is correlated with both preceding and subsequent acceleration bursts. a, **b**, Distribution of latencies from each significant calcium transient onset (mean whole-field fluorescence; 6 mice) to the first acceleration burst onset within 1 s preceding ($n = 1{,}087$, 6 mice, **a**) or following ($n = 990$, 6 mice, **b**) during continuous locomotion. Latencies are less variable (*F*-test for difference between variance of latencies, $P = 7.1 \times 10^{-5}$) and shorter (Wilcoxon test for difference between latency means, $P = 1.2 \times 10^{-5}$) to the preceding acceleration onsets, indicating more precise relative timing between the GCamp6f transients and the preceding acceleration burst versus the following acceleration burst. **c**, **d**, Mean acceleration traces from the first acceleration (within 1 s) preceding (**c**) or following (**d**) all short duration (<0.5 s) large amplitude (>75th percentile, $n = 149$ transients, grey) and small amplitude (<25th percentile, $n = 149$ transients, bronze) calcium transients occurring during continuous locomotion; aligned on acceleration onsets. Insets are schematics of the GCaMP6f transients. A significant correlation is present between the transient amplitudes and the immediately preceding acceleration amplitudes (Spearman's Rho = 0.16, $P = 1.2 \times 10^{-4}$, from all transient-acceleration pairs; binned data from this plot shown in **c**). A significant correlation is also present between the transient amplitudes and the immediately following acceleration amplitudes (Spearman's Rho = 0.13, $P = 0.006$, from all transient-acceleration pairs; binned data from this plot shown in **d**). **e**, Schematic summarizing relationship between the timing and amplitude of dopamine axon calcium transients and acceleration bursts during continuous locomotion. **f**, Mean acceleration (black) and whole-field $\Delta F/F$ (green) all triggered on all accelerations during continuous locomotion that were less than 1.7 m s^{-2} in amplitude ($n = 596$ accelerations, $n = 6$ mice); this demonstrates that dopamine axon GCaMP6f signalling displays a timing preference with respect to small amplitude accelerations, with similar timing and amplitude to that shown in Fig. 2b (which includes both large and small amplitude accelerations).



Extended Data Figure 6 | Pulsed optogenetic stimulation of dorsal-striatum-projecting and ventral-striatum-projecting dopamine axons. **a–d**, Pulsed optogenetic stimulation of dorsal-striatum-projecting dopamine axons can entrain accelerations during locomotion. **e–h**, Pulsed optogenetic stimulation of ventral-striatum-projecting dopamine axons leads to little effect on locomotion. **a**, **b**, Representative acceleration traces from continuous locomotion periods during (and initiated by) 6 Hz (**a**) and 3 Hz (**b**) laser stimulation trains in the same mouse. Blue, laser stimulation train, one mouse out of seven. **c**, **d**, Mean accelerations triggered on individual laser burst onsets during continuous locomotion periods for 6 Hz (**c**) and 3 Hz (**d**) across all laser bursts in all mice and sessions ($n = 7$ mice). **e**, Mean absolute value of mouse accelerations aligned on onset of laser stimulation train applied to mice at rest (mean across all stimulation onsets, $n = 55$ and 91 for ventral and dorsal respectively, in all sessions and mice, 3 and 6 Hz stimulation included). Dorsal and ventral striatum stimulations are from same ChR2-expressing mice ($n = 4$). Three mice were not stimulated in ventral striatum and thus not included in this figure. Mean acceleration elicited by ventral stimulation was significantly ($P < 0.01$, Wilcoxon rank-sum test) less than that elicited by dorsal axon stimulation. However, acceleration from ventral stimulation was significantly greater than chance ($P < 0.01$ shuffle test). This small effect in the ventral striatum could be due to activation of fibres, which also project to the dorsal striatum or to an increase in arousal. Although note that acceleration frequency during locomotion was not altered for stimulation in ventral striatum (see **f**, **h**). **f**, **g**, Centre of mass of acceleration power spectra for each mouse for locomotion periods initiated during 3 or 6 Hz stimulations ($n = 4$ mice). Horizontal bars indicate means, lines connect same mouse. **h**, Mean difference between the centre of mass of the acceleration power spectra computed for locomotion periods initiated during 3 Hz or 6 Hz axon stimulations in ventral (**f**) or dorsal (**g**) striatum. Positive values indicate a shift towards higher frequency accelerations for 6 Hz stimulations. $*P < 0.05$, Wilcoxon rank-sum test.

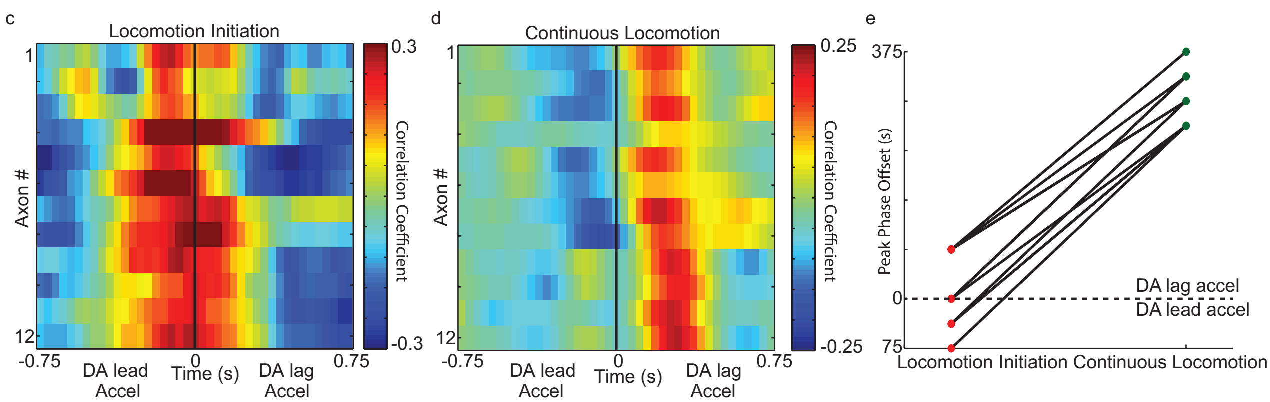


Extended Data Figure 7 | Histology and response distributions from each sparsely injected mouse. **a**, GCaMP6f expression (green) and immunofluorescence (red) from all VTA-targeted mice ($n = 5$). **b**, Reward response versus locomotion index (as in Fig. 4n) for each axon recorded from the corresponding mice in **a**. Green, significant locomotion; red, significant reward; blue both significant; neither significant not shown. **c**, **d**, Same as **a** and **b**, except from all SNc-injected mice ($n = 5$). Six out of ten mice were not stained for tyrosine hydroxylase. Scale bars, 500 μm.

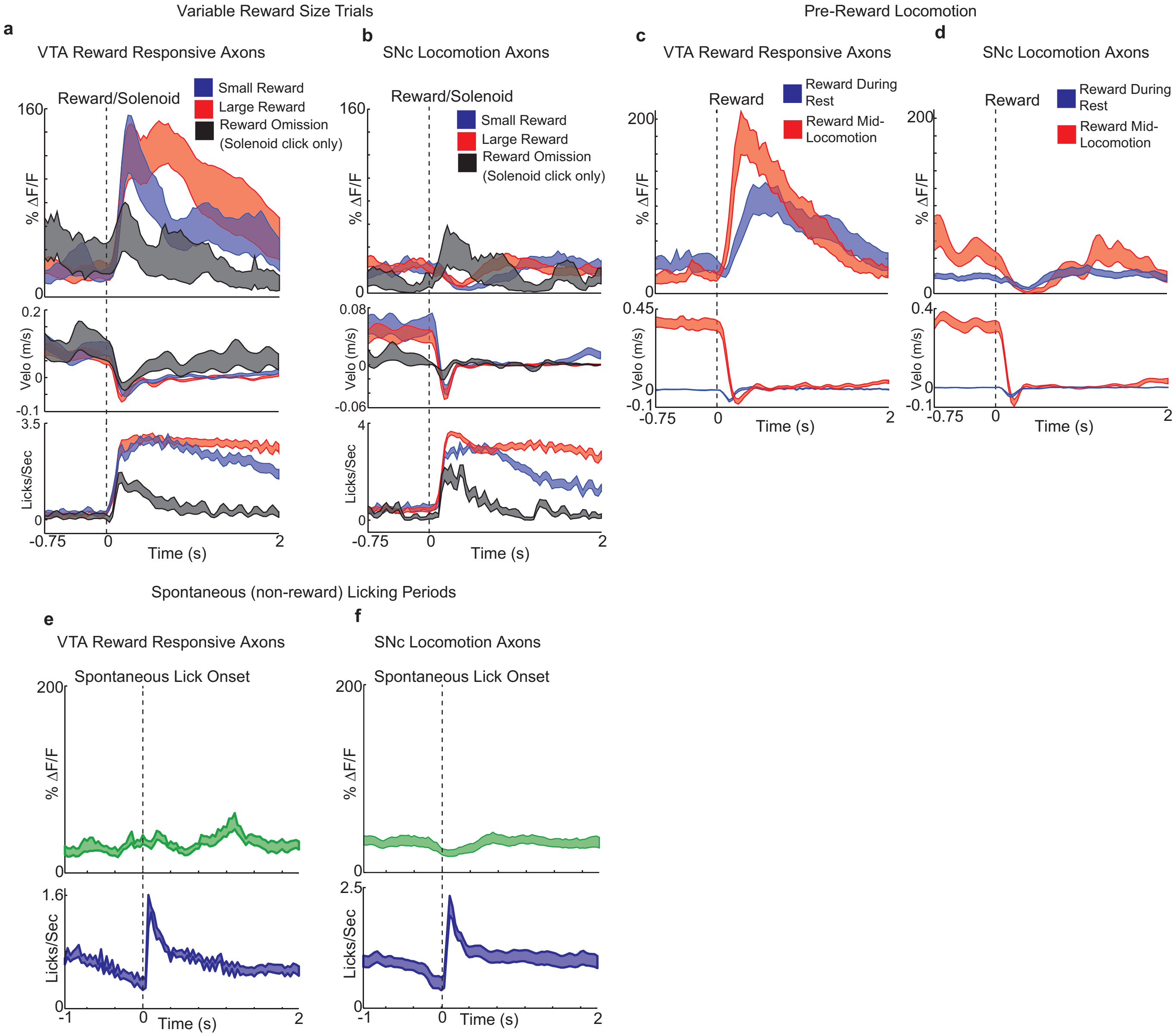


b

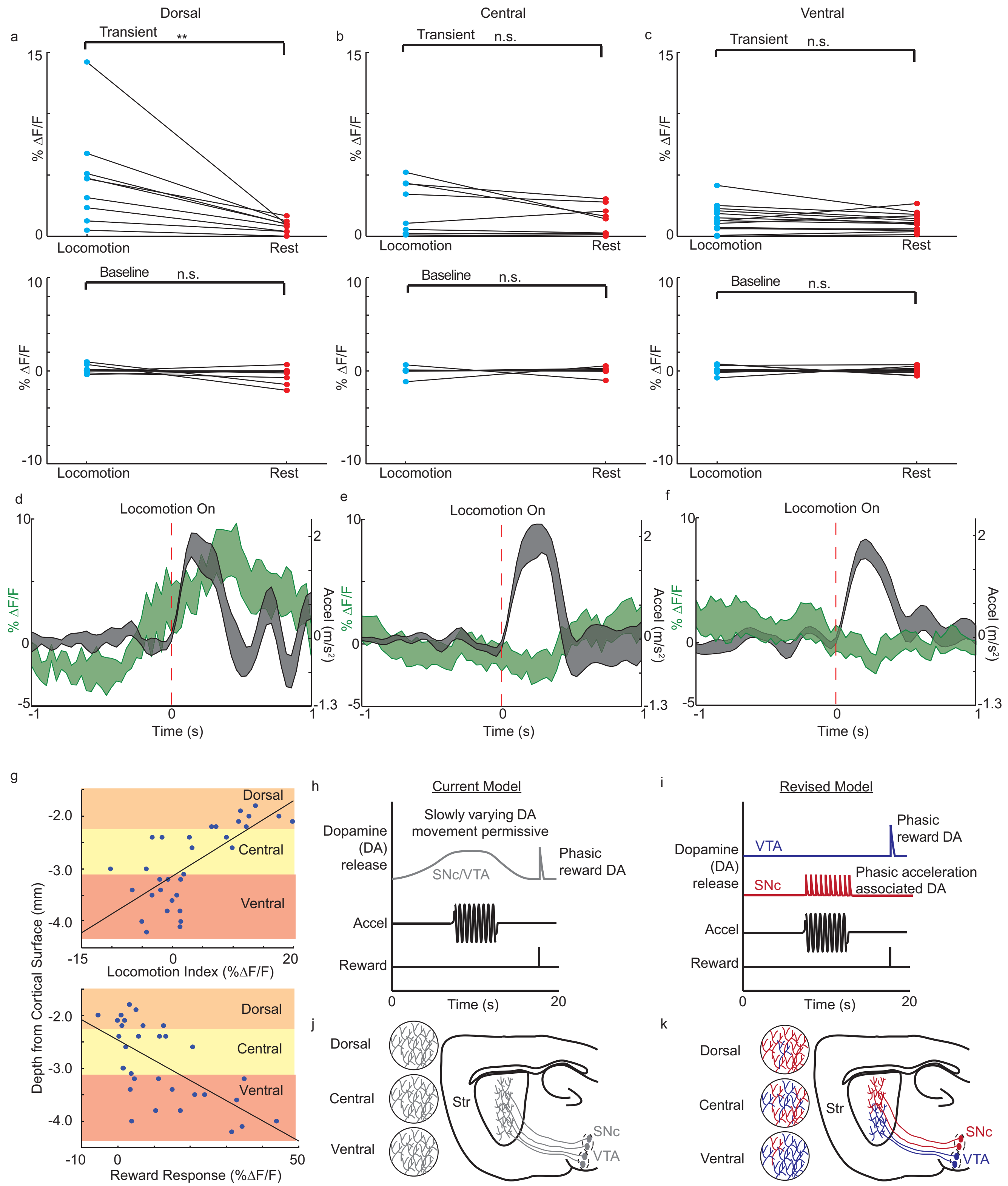
	Clustering Threshold	Number of locomotion only axons	% Locomotion only axons	Number of reward only axons	% Reward only axons	Number of axons correlated to both	% Both axons	Number of axons correlated to neither	% Neither axons	Total number of axons
SNc	No Clustering	95	83%	0	0%	0	0%	19	17%	114
	0.8	63	86%	0	0%	0	0%	10	14%	73
	0.5	40	87%	0	0%	0	0%	6	13%	46
	0.2	37	90%	0	0%	0	0%	4	10%	41
VTA	No Clustering	47	23%	96	47%	21	10%	42	20%	206
	0.8	19	19%	43	44%	10	10%	26	27%	98
	0.5	9	16%	25	45%	8	14%	14	25%	56
	0.2	8	16%	26	52%	7	14%	9	18%	50



Extended Data Figure 8 | Distribution of reward and locomotion indexes and fraction of reward and locomotion signalling axons from VTA and SNc are highly similar using different correlation thresholds for clustering axon segments (a, b); single axon signaling timing during locomotion onset versus continuous locomotion (c–e). a, Reward response versus locomotion index for putative single axons from SNc (*n* = 5 mice, top row) and VTA (*n* = 5 mice, bottom row) using different correlation thresholds (no clustering, 0.5, and 0.2) for hierarchical clustering of activity patterns (see Methods). Axons are colour-coded by significant responses to locomotion (green), reward (red), or both (blue). Note that despite the total number of putative axons decreasing with correlation threshold, the inverse relationship between locomotion and reward signalling across the population remains the same. **b**, Table showing the total numbers and fractions of responsive axons across the VTA and SNc populations for different clustering thresholds. Note that despite the total number of putative axons decreasing with correlation threshold, the fraction of axons signalling either reward, locomotion, both or neither is highly similar. **c**, Correlations (Pearson's) between acceleration and selected putative single SNc axon $\Delta F/F$ traces at different relative time-lags (that is, cross-correlations) during locomotion initiation periods; each row is mean for each axon for a single session (axons from *n* = 3 out of 5 mice). **d**, Same as **c**, but during continuous locomotion periods; same axons during same sessions as in **c**. **e**, Peak cross-correlation times for data shown in **c** and **d** (lines connect same axons during same sessions).



Extended Data Figure 9 | Further characterization of putative single dopamine axons in relation to reward and licking. a, Mean $\Delta F/F$ trace for VTA reward responsive axons (Methods, $n=23$ axons, $n=4$ mice with variable reward sessions) (top), velocity (mid), and licking (bottom) triggered on large volume (red), small volume (blue) and omission (black, solenoid click was present, but no reward delivered; Methods, $n=17$ axons, 3 mice with omission sessions) reward deliveries. **b**, Same as **a** except for SNc locomotion responsive axons ($n=62$ and 18 axons for reward and omission traces respectively). **c**, Mean VTA reward axon $\Delta F/F$ trace (top) and velocity (bottom) triggered on reward deliveries during continuous locomotion (red, $n=25$ axons) or rest periods (blue, $n=37$ axons). **d**, Same as **c** except for SNc locomotion responsive axons ($n=25$ and 62 axons for locomotion and rest respectively). **e**, Mean VTA reward axon $\Delta F/F$ (top) and mouse licking (bottom) triggered on spontaneous, non-reward licking onsets ($n=15$ axons, 3 mice). **f**, Same as **e** except for SNc locomotion responsive axons ($n=15$ axons, 3 mice). Mice that did not lick outside reward periods were excluded. Shaded regions in **a**–**f** denote mean ± s.e.m.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Dopamine axon locomotion signalling measured by fibre photometry from different striatal sub-regions. **a**, Top, comparison of mean photometry fluorescence $\Delta F/F$ (mean of significant transients, excluding baseline periods) recorded from dorsal striatum between locomotion and resting periods; each point represents mean $\Delta F/F$ for running or resting over one session for recording from a single dorsal striatum location (lines connect same recording location/session; $n = 5$ mice). Bottom, comparison of mean baseline (periods with no significant calcium transients) photometry $\Delta F/F$ recorded from dorsal striatum between locomotion and resting periods; each point represents mean baseline $\Delta F/F$ for running or resting over one session for recording from a single dorsal striatum location (lines connect same recording location/session, $n = 5$ mice). **b**, **c**, Same as **a**, except for recordings from central and ventral striatum, respectively. **d**, Mean photometry $\Delta F/F$ recorded from dorsal striatum triggered on locomotion initiations (mean across all initiations, $n = 5$ mice). **e**, **f**, Same as **d**, except for recordings from central and ventral striatum, respectively. **g**, Locomotion index (top) and reward response (bottom) versus striatum recording depth (from data presented in Fig. 5). **h**, Schematic of prominent current model for dopamine signalling dynamics in the striatum. **i**, Schematic of our new model for dopamine signalling dynamics based on data presented here. **j**, **k**, Saggital schematics illustrating current homogenous dopamine signalling model (**j**) and our new model incorporating functional heterogeneity (**k**). $^{**}P < 0.01$, Wilcoxon rank-sum test; n.s., not significant. Shaded regions in **d**–**f**, mean ± s.e.m across initiations ($n = 20$, 28 and 56 from dorsal, central and ventral respectively in 5 mice).

doi:10.1038/nature18634

Human commensals producing a novel antibiotic impair pathogen colonization

Alexander Zipperer[1,2]*, Martin C. Konnerth[3]*, Claudia Laux[1,2], Anne Berscheid[4], Daniela Janek[1,2]†, Christopher Weidenmaier[2,5], Marc Burian[6], Nadine A. Schilling[3,7], Christoph Slavetinsky[1,2], Matthias Marschal[5], Matthias Willmann[2,5], Hubert Kalbacher[7], Birgit Schittek[6], Heike Brötz-Oesterhelt[2,4], Stephanie Grond[3], Andreas Peschel[1,2] & Bernhard Krismer[1,2]

The vast majority of systemic bacterial infections are caused by facultative, often antibiotic-resistant, pathogens colonizing human body surfaces. Nasal carriage of *Staphylococcus aureus* predisposes to invasive infection, but the mechanisms that permit or interfere with pathogen colonization are largely unknown. Whereas soil microbes are known to compete by production of antibiotics, such processes have rarely been reported for human microbiota. We show that nasal *Staphylococcus lugdunensis* strains produce lugdunin, a novel thiazolidine-containing cyclic peptide antibiotic that prohibits colonization by *S. aureus*, and a rare example of a non-ribosomally synthesized bioactive compound from human-associated bacteria. Lugdunin is bactericidal against major pathogens, effective in animal models, and not prone to causing development of resistance in *S. aureus*. Notably, human nasal colonization by *S. lugdunensis* was associated with a significantly reduced *S. aureus* carriage rate, suggesting that lugdunin or lugdunin-producing commensal bacteria could be valuable for preventing staphylococcal infections. Moreover, human microbiota should be considered as a source for new antibiotics.

Infections caused by highly antibiotic-resistant bacteria have greatly increased in recent years and represent a major cause of morbidity and mortality worldwide, including in developed countries[1,2]. Multidrug resistant organisms (MDRO), such as methicillin-resistant *S. aureus*[3], vancomycin-resistant enterococci[4], and third-generation cephalosporin-resistant Gram-negative bacteria[5], are expected to become more frequent causes of death than cancer in the coming decades[6]. Despite the urgent need for new antibiotics that are effective against resistant bacteria, very few compounds are in development, the majority of which are congeners of currently used antibiotic classes[7,8]. Nevertheless, innovative approaches for cultivation of yet uncultured potential antibiotic producers or activation of silent biosynthetic gene clusters have recently led to the discovery of entirely new antimicrobial compounds with useful properties[9–12]. Moreover, new compound sources, such as the large inventory of antimicrobial host defence peptides from higher organisms, open new avenues for the development of new anti-infectives[13,14]. As the available antibiotics lose their efficacy, it is important to limit the ongoing spread of resistant bacteria[15]. Unfortunately, most of the major MDROs are increasingly disseminating in the community and cannot be effectively targeted by standard infection-control measures[2]. The strong antibiotic selection pressure in humans and in farm animals and the increasing fitness of methicillin-resistant *S. aureus*, vancomycin-resistant enterococci, and third-generation cephalosporin-resistant Gram-negative bacteria in competition with commensals leads to the symptomless and usually unrecognized presence of MDRO in the microbiota of many humans[16–18]. Notably, the vast majority of systemic bacterial infections are caused by endogenous pathogens from human microbiota, and MDRO-colonized individuals are exposed to a substantially higher risk of invasive infections, that are difficult to treat, when they undergo surgery or immunosuppression, or suffer from trauma[1,19,20]. *S. aureus* is found in the anterior nares of approximately 30% of the human population. Eradication by the antibiotic mupirocin strongly reduces predisposition to invasive infection[19]. However, mupirocin has to be applied repeatedly over five days, requiring strict compliance and costly extension of patient hospitalization; in addition, rates of mupirocin resistance are steadily increasing[21]. Moreover, current decolonization strategies targeting MDRO at other body surfaces, such as the intestine, depend on broad-spectrum antibiotics and are highly controversial[22]. Thus, there is not only an urgent need for new therapeutic antibiotics but also for new effective MDRO decolonization strategies.

Although many ongoing research efforts address major virulence and resistance mechanisms, the processes governing bacterial fitness, competition, and dissemination in microbiota remain poorly understood. Recent studies have shown how diverse and dynamic human microbiota are, especially those of the environment-exposed microbiota of the skin[23] and upper airways[24]. These ecological niches are particularly poor in nutrients[25], suggesting that colonizing bacteria are probably in strong competition and may use a variety of strategies to overcome competitors[25,26]. Bacteria from human microbiota have occasionally been found to produce bacteriocins, antimicrobial substances acting against closely related bacteria[27,28]. Genes related to antibiotic biosynthesis have been identified in human metagenomes[29] leading to the discovery of lactocillin, a ribosomally synthesized thiopeptide antibiotic produced by a human vaginal commensal[30]. However, the potential roles of such complex bioactive compounds in shaping human microbiota are still unknown.

Antimicrobial activity of a nasal commensal

A previously described collection of nasal *Staphylococcus* isolates[25] was screened for antimicrobial activity against *S. aureus*. Whereas most

[1]Interfaculty Institute of Microbiology and Infection Medicine, Infection Biology, University of Tübingen, 72076 Tübingen, Germany. [2]German Centre for Infection Research (DZIF), Partner Site Tübingen, 72076 Tübingen, Germany. [3]Institute of Organic Chemistry, University of Tübingen, 72076 Tübingen, Germany. [4]Interfaculty Institute of Microbiology and Infection Medicine, Microbial Bioactive Compounds, University of Tübingen, 72076 Tübingen, Germany. [5]Interfaculty Institute of Microbiology and Infection Medicine, Medical Microbiology, University of Tübingen, 72076 Tübingen, Germany. [6]Department of Dermatology, Division of Dermatooncology, University of Tübingen, 72076 Tübingen, Germany. [7]Interfaculty Institute of Biochemistry, University of Tübingen, 72076 Tübingen, Germany. †Present address: Boehringer Ingelheim, 88400 Biberach, Germany.
*These authors contributed equally to this work.

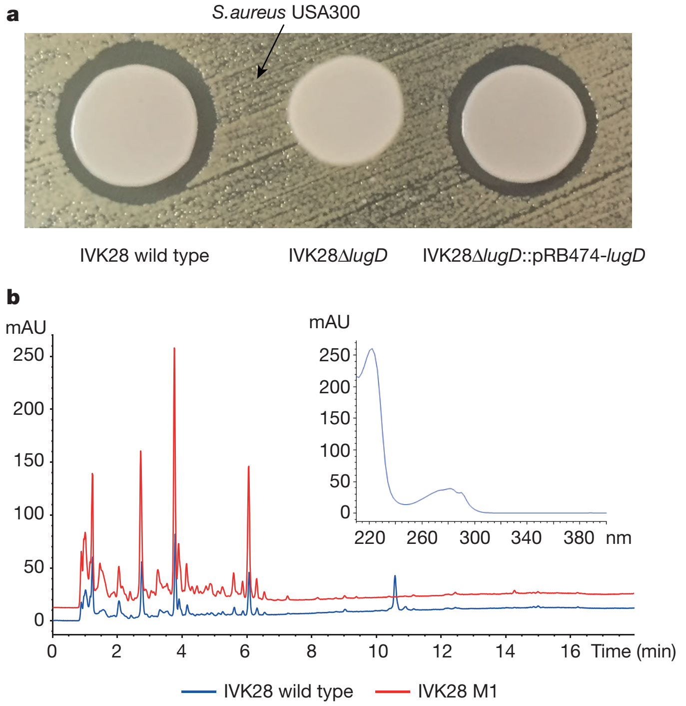


Figure 1 | Lugdunin production by wild-type *S. lugdunensis* and isogenic mutants. a, Bioactivity assay with the *S. lugdunensis* IVK28 wild type, the Δ*lugD* mutant and the complemented mutant against *S. aureus* USA300. Basic medium plates were inoculated with a lawn of *S. aureus* USA300. *S. lugdunensis* IVK28 cells from overnight cultures of wild type, the lugdunin-deficient deletion mutant Δ*lugD* and complemented mutant Δ*lugD*::pRB474-*lugD* were spotted onto the plates. **b**, HPLC-UV-chromatogram shows comparison of *S. lugdunensis* IVK28 wild type and the lugdunin-deficient transposon mutant M1. Cell extracts of the IVK28 wild type (blue) and the transposon mutant M1 (red) were compared by reversed-phase HPLC-UV. The inlay depicts the absorption spectrum of lugdunin with the prominent absorption at 280 nm, indicating the presence of tryptophan. Absorbance was measured in milli absorption units (mAU).

isolates did not affect *S. aureus*, the strain *S. lugdunensis* IVK28 was found to have a particularly strong capacity to prevent the growth of *S. aureus* (Fig. 1a). IVK28 produced an antibacterial substance only under iron-limiting conditions and only on solid agar surfaces, not in liquid culture.

Transposon mutagenesis of IVK28 led to isolation of mutant M1, which did not inhibit *S. aureus*. Analysis of the transposon insertion site revealed the disruption of an uncharacterized gene encoding a putative non-ribosomal peptide synthetase (NRPS; position 860375/76 of SLUG_RS03940 in the annotated genome sequence of *S. lugdunensis* N920143 ref. 31; accession number, NC_017353.1). This gene is encoded in an operon of approximately 30 kbp with several other NRPS and antibiotic biosynthesis-related genes (Fig. 2a and Extended Data Fig. 1a), suggesting that the inhibitory molecule of IVK28 may be a complex, non-ribosomally synthesized peptide compound. The NRPS operon was found in all *S. lugdunensis* genomes in the databases, and it was identified by PCR in each of the *S. lugdunensis* strains from our strain collection, indicating that it is characteristic of the species, rather than a strain-specific feature. Nevertheless, the GC-content of the operon of 26.9% is clearly distinct from that of the whole genome (33.8%) indicating that the gene cluster may have been transferred to *S. lugdunensis* by horizontal gene transfer from another bacterial species. The operon consists of four NRPS genes (named *lugA, B, C, D*) encoding adenylation domains for five amino acids (Fig. 2b). The *lug* operon further includes all genes whose products are required for the synthesis and transport of a non-ribosomally synthesized peptide compound (Extended Data Fig. 1a). Gene clusters for secondary metabolite production are frequently found in streptomycetes and other soil bacteria but are rare in human-associated bacteria[29]. The *lug* operon was exclusively found in *S. lugdunensis* and encodes a unique combination of antibiotic biosynthesis enzymes, all with less than 35% identity to any other described enzyme, suggesting that it may be responsible for biosynthesis of a novel compound. To confirm that the *lug* operon is responsible for the antimicrobial activity of IVK28, the smallest NRPS gene, *lugD*, was deleted by gene replacement. The mutant Δ*lugD* showed no detectable antimicrobial activity, but the phenotype was restored by complementation with a plasmid-encoded copy of *lugD* (Fig. 1a).

Lugdunin, a novel peptide antibiotic

The antimicrobial activity of IVK28 was enriched by ethanol extraction of agar-grown cells. Reversed-phase high-pressure liquid chromatography (HPLC) coupled with ultraviolet (UV) and mass spectrometry revealed obvious differences between extracts from IVK28 wild-type and mutant M1 in only one signal (Fig. 1b). The molecular formula, $C_{40}H_{62}N_8O_6S$, deduced from electrospray ionization high-resolution mass spectrometry (Extended Data Fig. 2a, b; $[M+H]^+$ $C_{40}H_{63}N_8O_6S^+$, calculated 783.45858, found 783.45850, Δ0.1 parts per million (p.p.m.), relative molecular mass = 783.03) did not correspond to any known molecule. Spectral UV absorbance at 280 nm indicated the presence of a tryptophan moiety (Fig. 1b). As the compound was not produced in liquid culture, which impeded isolation of sufficient amounts for chemical characterization and biological profiling, expression of the biosynthetic genes was uncoupled from the native *lug* operon promoter and its regulation. This strategy has recently been successfully applied to other antimicrobial gene clusters[10,32,33]. The putative *tetR*-like regulator gene of the *lug* operon (*lugR*) was deleted, and the strong *xylAB* promoter was inserted upstream of the NRPS genes, along with the gene for its corresponding xylose-sensitive repressor XylR. Cultivation of the resulting strain *S. lugdunensis* IVK28-Xyl (Extended Data Fig. 1b) in liquid culture with xylose led to strong production of inhibitory activity, which could be purified by consecutive steps of 1-butanol extraction, gel filtration, and reversed-phase HPLC. The pure compound exhibited exactly the same mass as determined for the compound from the wild-type strain IVK28 and was named lugdunin.

Nuclear-magnetic resonance (NMR), electrospray ionization high-resolution mass spectrometry, and an advanced Marfey's analysis of lugdunin revealed a cyclic peptide (Fig. 2c, Extended Data Fig. 3a–e, and Extended Data Table 1) comprising an unusual thiazolidine heterocycle and five amino acids (D-valine, L-tryptophan, D-leucine, L-valine, and D-valine) (Fig. 2c). The thiazolidine building block occurs in certain linear NRPS products, such as watasemycins[34] and yersiniabactin[35], but is yet unreported in macrocyclic peptides (Extended Data Fig. 4). The lugdunin thiazolidine ring is probably formed by condensation of an N-terminal L-cysteine with a C-terminal L-valine residue (position one and seven, respectively, of the precursor peptide) upon reductive release of a linear heptapeptide aldehyde from the NRPS mega-enzyme by the terminal reductase of LugC (Fig. 2b and Extended Data Fig. 4d). Total chemical synthesis of lugdunin yielded a product with identical chemical properties and antimicrobial activity compared to the natural product from strain IVK28, which confirmed the assigned lugdunin structure.

The origin of lugdunin from a heptapeptide was unexpected because the NRPS proteins from the *lug* operon encompass only five adenylation domains, central enzymatic domains whose number usually determines the number of incorporated amino acids (Fig. 2b). The predicted specificities of NRPS adenylation domains[36] were in agreement with the identified amino acids, except for position three (tryptophan instead of threonine), indicating a new specificity for the second adenylation domain of LugA. Moreover, LugC exhibited an unusual modular organization[37] with one adenylation domain (valine) but two downstream peptide-bond-forming condensation and three peptidyl carrier protein domains required for amino acid transfer. This

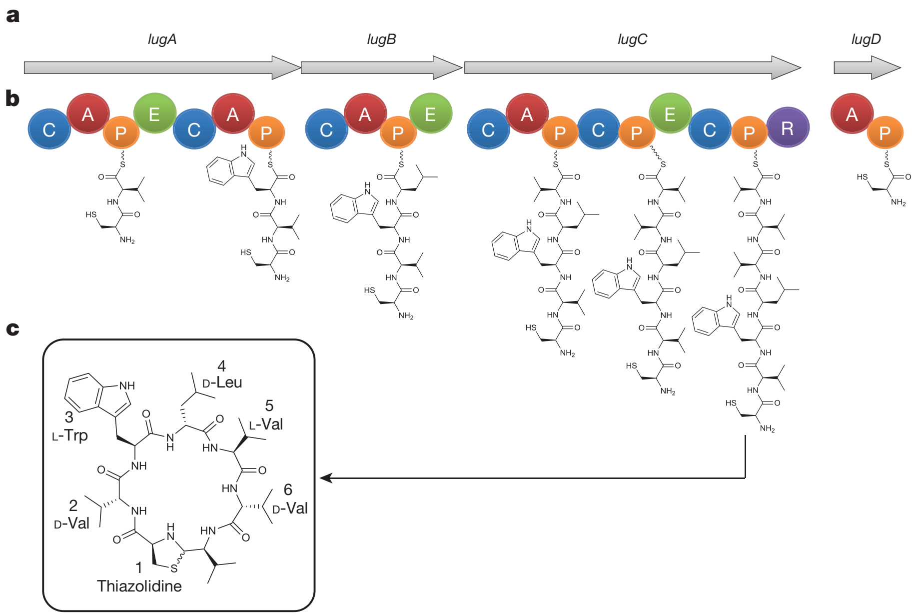


Figure 2 | Gene cluster, proposed biosynthesis pathway and chemical structure of lugdunin. a, NRPS genes *lugA*, *B*, *C*, *D*. **b**, Modular organization of gene products. Functional domains: A, adenylation; P, peptidyl carrier protein; C, condensation; E, epimerization; R, reductase. The sequential biosynthesis of lugdunin starts presumably at the characteristic initiation module of LugD and continues with LugA–C. The C-terminal reductase domain of LugC catalyses aldehyde formation at the thioester of the terminal valine and enables subsequent ring closure of the thiazolidine, leading to the cyclopeptide. **c**, Chemical structure of lugdunin. Cyclization of the reactive C-terminal aldehyde and N-terminal cysteine is assumed to occur via thiohemiacetal to the thiazolidine building block (Extended Data Fig. 4d).

organization suggested that the single adenylation domain of LugC is responsible for activation of three consecutive valine units, which are subsequently incorporated in alternating L- and D-configurations. A partially similar mechanism has been described for yersiniabactin biosynthesis[38] (see also Supplementary Discussion). Thus, the LugC domains represent an unusual assembly line compared to characterized NRPS machineries.

Activity against major human pathogens

Lugdunin has potent antimicrobial activity against a wide range of Gram-positive bacteria, including opportunistic pathogens such as difficult-to-treat methicillin-resistant *S. aureus*, glycopeptide-intermediate resistant *S. aureus* and vancomycin-resistant *Enterococcus* isolates. Minimal inhibitory concentration (MIC) values in the micromolar range (1.5–12 μg ml^{-1}; 1.9–15.3 μM) demonstrated high potency, and this activity was not impaired in the presence of human serum (Table 1). Lugdunin was bactericidal against methicillin-resistant *S. aureus* with complete killing at 10 × MIC (Fig. 3a). Lugdunin did not cause lysis of primary human neutrophils or erythrocytes, and even high amounts of lugdunin showed no substantial inhibition of the metabolic activity of the human monocytic cell line HL60 ($IC_{50} > 50$ μg ml^{-1}) (Extended Data Fig. 5). Bacterial cells exposed to lugdunin stopped incorporating radioactive DNA, RNA, protein or cell-wall precursors almost simultaneously even at concentrations below the MIC (Extended Data Fig. 6), suggesting that lugdunin may lead to rapid breakdown of bacterial energy resources. In this respect, lugdunin resembles daptomycin, which has been shown to cause parallel cessation of all four metabolic pathways[39] and whose exact mode of action is still unknown. During lugdunin purification, a 2-D-*allo*-isoleucine variant (instead of 2-D-valine) was isolated. This derivative with an additional methyl group and stereogenic centre was produced by *S. lugdunensis* in small amounts and was inactive. Development of resistance was not observed in *S. aureus* during continuous serial passaging in the presence of subinhibitory concentrations of lugdunin over 30 days (Fig. 3b). In contrast, *S. aureus* rapidly developed resistance to rifampicin within a few days of exposure (Fig. 3b).

The capacity of lugdunin to cure infections *in vivo* was analysed in a mouse skin infection model reflecting a typical human *S. aureus* infection and a common indication for antibiotic treatment[40,41]. The back skin of shaved black-6 mice (C57BL/6) was superficially damaged by multiple stripping with adhesive tape and was infected with *S. aureus*. Then, 24, 30, and 42 h after infection, mice were treated with 1.5 μg lugdunin per time point and infection site, and 6 h later mice were euthanized. Lugdunin treatment led to a strong reduction or even complete eradication of viable *S. aureus* on the surface and in the deeper layers of the skin (Figs 3c, d), demonstrating that lugdunin eradicates *S. aureus* and penetrates tissues *in vivo*. Note that infection was not affected by lugdunin in two samples, most probably because some animals removed the lugdunin ointment by intensive licking. Bacteria obtained from these skin samples exhibited unchanged susceptibility to lugdunin, indicating that they had not developed resistance.

Table 1 | Lugdunin spectrum of activity

Species and strain	Resistance	Lugdunin MIC (μg ml^{-1})
Staphylococcus aureus USA300 (LAC)	MRSA	1.5
+ 50% human serum		1.5
Staphylococcus aureus USA300 (NRS384)	MRSA	1.5
Staphylococcus aureus Mu50	GISA	3
Staphylococcus aureus SA113		3
Staphylococcus aureus RN4220		3
Enterococcus faecium BK463	VRE	3
Enterococcus faecalis VRE366	VRE	12
Listeria monocytogenes ATCC19118		6
Streptococcus pneumoniae ATCC49619		1.5
Bacillus subtilis 168 (*trpC2*)		4
Pseudomonas aeruginosa PAO1		>50
Escherichia coli DH5α		>50

MRSA, methicillin-resistant *S. aureus*; GISA, glycopeptide intermediate-resistant *S. aureus*; VRE, vancomycin-resistant *Enterococcus*.

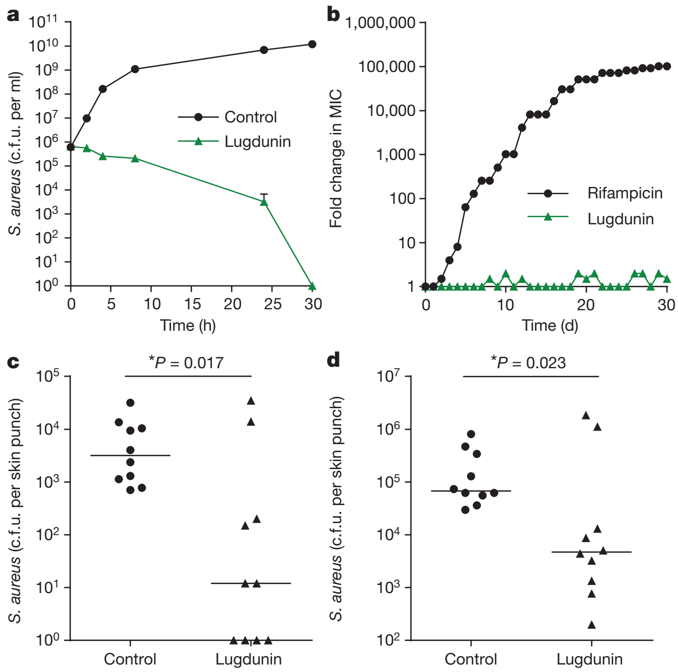


Figure 3 | Lugdunin has bactericidal activity, a low risk for resistance and efficacy in a mouse skin infection model. a, Killing curve. Incubation of *S. aureus* with a 10 × MIC of lugdunin leads to complete killing of the inoculum after 30 h (detection limit is 10° colony-forming units (c.f.u.) per ml). Data represent medians ± s.d. of three independent experiments. **b**, Serial passaging of *S. aureus* with sub-inhibitory concentrations of rifampicin leads to high resistance against rifampicin. However, development of resistance is not observed with lugdunin. A representative of two independent experiments is shown. **c**, **d**, Mouse skin infection. Lugdunin treatment (3 times, 1.5 μg per skin area) of *S. aureus* skin infections (5 mice corresponds to 10 skin punches per group) leads to strongly reduced numbers of viable bacteria after two days. c.f.u. for surface-attached bacteria (**c**) or bacteria located in the deeper skin tissue (**d**) per skin punch are shown. Horizontal lines represent the median of each group. Significant differences between groups were analysed by the Mann–Whitney test (*$P < 0.05$).

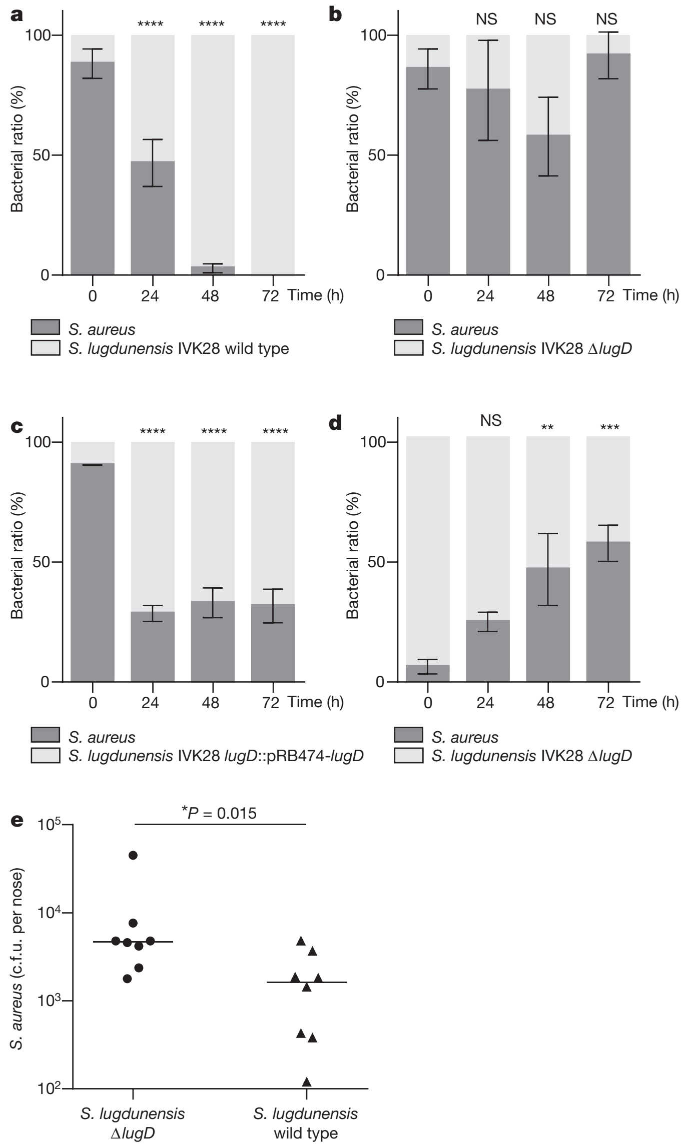


Figure 4 | Lugdunin-producing wild-type *S. lugdunensis* IVK28 restricts the growth of *S. aureus in vitro* and *in vivo* in cotton rats. a, *S. aureus* is overgrown by wild-type *S. lugdunensis* IVK28 (dark or light grey columns, respectively) on agar plates inoculated at ratios of ~90:10. **b**, In contrast, the ratio between IVK28 Δ*lugD* and *S. aureus*, inoculated at a ratio of ~90:10, does not change significantly over time. **c**, The capacity of *S. lugdunensis* Δ*lugD* to overgrow *S. aureus* is largely restored by plasmid-encoded *lugD*. **d**, *S. aureus* displaces IVK28 Δ*lugD* even when inoculated at ratios of ~10:90. Data represent mean values ± s.d. of three independent experiments. Significant differences between starting conditions and indicated time points were analysed by one-way ANOVA (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; NS, not significant). **e**, Cotton rat noses (8 animals per group) co-colonized by *S. aureus* and *S. lugdunensis* IVK28 wild type show significantly less *S. aureus* c.f.u. after five days compared to *S. lugdunensis* IVK28 Δ*lugD*. Horizontal lines represent the median of each group. Significant differences, calculated by the Mann–Whitney test, are indicated (*$P < 0.05$).

Lugdunin production outcompetes *S. aureus*

The production of antimicrobials, mostly plasmid-encoded ribosomally synthesized bacteriocins, has been sporadically documented in individual bacterial strains from human microbiota[27]. However, the roles of such compounds in microbial fitness and in microbiota dynamics have remained largely unknown. To determine whether lugdunin contributes to the capacity of *S. lugdunensis* IVK28 to prevail in competition with *S. aureus*, the two species were co-cultivated on solid agar surface, promoting lugdunin production, and bacterial numbers were monitored for three days.

As shown in Fig. 4a, the lugdunin-producing IVK28 wild type overgrew *S. aureus* efficiently, even when the inoculum contained tenfold more *S. aureus* than *S. lugdunensis* cells. No viable *S. aureus* cells were recovered after three days, indicating complete killing by *S. lugdunensis*. In contrast, IVK28 Δ*lugD* could not outcompete *S. aureus* (Fig. 4b) and was even overgrown when inoculated at tenfold higher numbers than *S. aureus* (Fig. 4d). The *S. aureus*-eradicating capacity of Δ*lugD* could be largely restored by complementation with *lugD* on a plasmid (Fig. 4c). These data demonstrate that *S. lugdunensis* IVK28 wild type can effectively eradicate *S. aureus* and that lugdunin production is responsible for this trait.

Nasal carriage is known to be a major risk factor for invasive *S. aureus* infections[19,42]. To explore whether *S. lugdunensis* can interfere with nasal *S. aureus* colonization *in vivo* in vertebrates, the noses of cotton rats, a well-established animal model for investigating *S. aureus* nasal colonization[43], were instilled with mixtures of *S. lugdunensis* IVK28 wild type or Δ*lugD* plus *S. aureus*. The three test strains colonized cotton rat noses stably over the 5-day test period when instilled individually (Extended Data Fig. 7). However, when the two species were co-inoculated, significantly less *S. aureus* cells were retrieved from animals co-colonized by IVK28 wild type compared to those co-colonized with Δ*lugD* (Fig. 4e). This finding indicates that lugdunin production can effectively interfere with *S. aureus* colonization *in vivo*.

Table 2 | *S. aureus* and *S. lugdunensis* distribution in hospitalized patients

	S. lugdunensis-positive	*S. lugdunensis*-negative	Total
S. aureus-positive	1	59	60
S. aureus-negative	16	111	127
Total	17	170	187
Risk§	0.059*	0.347*	0.321

Significant differences between *S. lugdunensis*-positive and *S. lugdunensis*-negative patients were analysed by the Chi-squared test.
§0.059 versus 0.347: risk ratio, 0.169; 95% confidence interval 0.025–1.147.
*$P = 0.015$.

Interference with *S. aureus* carriage

Factors governing human nasal *S. aureus* carrier status have remained largely unknown[42,44]. To investigate whether the presence of *S. lugdunensis* in the human nose can prevent co-colonization by *S. aureus*, we examined nasal swabs from 187 hospitalized patients for colonization by *S. lugdunensis, S. aureus* or both. The overall colonization rates of *S. aureus* and *S. lugdunensis* were 32.1% and 9.1%, respectively (Table 2), which corresponds to published data[20,45]. Of note, the percentage of *S. aureus* colonization in *S. lugdunensis*-positive individuals (5.9%) was approximately 5.9-fold lower than in *S. lugdunensis*-negative individuals (34.7%). Chi-squared statistical analysis showed that the reduced *S. aureus* detection rate in the presence of *S. lugdunensis* for the entire study population was significant (risk ratio 0.169; 95% confidence interval 0.025–1.147; $P = 0.015$; Table 2) indicating that strong interference precludes the simultaneous presence of *S. aureus* and *S. lugdunensis* in the human nose. Consistent with this finding, PCR analysis demonstrated that all nasal *S. lugdunensis* isolates contained the *lug* operon and all of the tested *S. aureus* strains (30 nasal as well as 7 laboratory) were found to be highly susceptible to lugdunin. Together with the ability of the lugdunin producer IVK28 to reduce *S. aureus* numbers in cotton rat noses, these data provide strong evidence for the potent capacity of *S. lugdunensis* to prevent human nasal colonization by *S. aureus*.

Conclusions

The human predisposition to *S. aureus* nasal carriage is governed by several host genetic or microbiota-related factors, which may affect the capacity of *S. aureus* to adhere to and multiply on nasal epithelia[43,44,46]. Here we provide evidence for a crucial role of *S. lugdunensis* and its antimicrobial product, lugdunin, in preventing *S. aureus* colonization of the human nose. *S. aureus* can only rarely be isolated from the axillae or groin, also habitats of *S. lugdunensis* in addition to nares[45], which may therefore be related to the production of lugdunin at these sites. Accordingly, utilizing lugdunin or lugdunin-producing commensal staphylococci could become a valuable strategy for preventing *S. aureus* colonization and invasive infections in high-risk patients, such as those undergoing elective surgery, immunosuppression, or regular haemodialysis. The concept of probiotic bacteria interfering with predisposition to infection has been pursued mostly for enteric pathogens[47]. Our study suggests that the probiotics concept should be extended to body sites other than the gut, such as the nasal mucous membranes, and to additional major pathogens such as *S. aureus*. *S. lugdunensis* is known as a rare cause of opportunistic infections[48], but a mutant lacking all potential virulence factors or introduction of the *lug* genes into an exclusively commensal species might enable the development of a safe probiotic strain.

Lugdunin represents the first example, to our knowledge, of a new class of compounds that we suggest should be named macrocyclic thiazolidine peptide antibiotics. Lugdunin shows bactericidal activity against many major pathogens combined with a particularly high barrier to resistance by mutation, even under prolonged selective pressure. The fact that all tested nasal and clinical *S. aureus* isolates maintained pronounced susceptibility to lugdunin suggests that it may be particularly difficult for bacteria to develop resistance to lugdunin. This indicates that lugdunin has apparently evolved for the purpose of bacterial elimination in the human organism, implying that it is optimized for efficacy and tolerance at its physiological site of action. Thus, lugdunin shows promise as a potential drug for inhibiting growth of *S. aureus* in the nares and potentially other body sites.

The genetic inventory of human-associated microbiomes is of great importance for many critical human body functions for example, degradation of ingested polymers, detoxification of xenobiotics, and release of immunomodulatory molecules[49]. Our discovery of commensal bacteria producing a potent antibiotic is in accord with recent reports on the presence of gene clusters for complex secondary metabolites in human-associated metagenomes[30] and suggests that many humans are constantly exposed to potent bioactive compounds from microbiota. The effect of such compounds on human body functions may vary substantially with the dynamic changes in microbiota structure. Whereas it has become increasingly difficult to identify novel compound structures from soil microbes such as actinomycetes, bacteria from human microbiota may become a valuable source for new types of antibiotics. It will be a challenge for future research to elucidate the identity, variability, activity, and ecological roles of such compounds and to exploit them for the development of new drugs.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 November 2015; accepted 9 June 2016.

1. Arias, C. A. & Murray, B. E. Antibiotic-resistant bugs in the 21st century—a clinical super-challenge. *N. Engl. J. Med.* **360,** 439–443 (2009).
2. Laxminarayan, R. *et al.* Antibiotic resistance—the need for global solutions. *Lancet Infect. Dis.* **13,** 1057–1098 (2013).
3. DeLeo, F. R., Otto, M., Kreiswirth, B. N. & Chambers, H. F. Community-associated meticillin-resistant *Staphylococcus aureus*. *Lancet* **375,** 1557–1568 (2010).
4. Arias, C. A. & Murray, B. E. The rise of the *Enterococcus*: beyond vancomycin resistance. *Nat. Rev. Microbiol.* **10,** 266–278 (2012).
5. Boucher, H. W. *et al.* Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **48,** 1–12 (2009).
6. WHO. Antimicrobial resistance: global report on surveillance 2014. http://www.who.int/drugresistance/documents/surveillancereport/en/ (2014).
7. Cooper, M. A. & Shlaes, D. Fix the antibiotics pipeline. *Nature* **472,** 32 (2011).
8. Bierbaum, G. & Sahl, H. G. The search for new anti-infective drugs: untapped resources and strategies. *Int. J. Med. Microbiol.* **304,** 1–2 (2014).
9. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517,** 455–459 (2015).
10. Laureti, L. *et al.* Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl Acad. Sci. USA* **108,** 6258–6263 (2011).
11. Hosaka, T. *et al.* Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **27,** 462–464 (2009).
12. Lincke, T., Behnken, S., Ishida, K., Roth, M. & Hertweck, C. Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angew. Chem. Int. Ed. Engl.* **49,** 2011–2013 (2010).
13. Schroeder, B. O. *et al.* Reduction of disulphide bonds unmasks potent antimicrobial activity of human β-defensin 1. *Nature* **469,** 419–423 (2011).
14. Zasloff, M. Inducing endogenous antimicrobial peptides to battle infections. *Proc. Natl Acad. Sci. USA* **103,** 8913–8914 (2006).
15. Bush, K. *et al.* Tackling antibiotic resistance. *Nat. Rev. Microbiol.* **9,** 894–896 (2011).
16. Marshall, B. M. & Levy, S. B. Food animals and antimicrobials: impacts on human health. *Clin. Microbiol. Rev.* **24,** 718–733 (2011).
17. Penders, J., Stobberingh, E. E., Savelkoul, P. H. & Wolffs, P. F. The human microbiome as a reservoir of antimicrobial resistance. *Front. Microbiol.* **4,** 87 (2013).
18. Davis, M. F., Price, L. B., Liu, C. M. & Silbergeld, E. K. An ecological perspective on U.S. industrial poultry production: the role of anthropogenic ecosystems on the emergence of drug-resistant bacteria from agricultural environments. *Curr. Opin. Microbiol.* **14,** 244–250 (2011).
19. Bode, L. G. *et al.* Preventing surgical-site infections in nasal carriers of *Staphylococcus aureus*. *N. Engl. J. Med.* **362,** 9–17 (2010).
20. Wertheim, H. F. *et al.* The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infect. Dis.* **5,** 751–762 (2005).
21. Thomas, C. M., Hothersall, J., Willis, C. L. & Simpson, T. J. Resistance to and synthesis of the antibiotic mupirocin. *Nat. Rev. Microbiol.* **8,** 281–289 (2010).

22. van der Meer, J. W. & Vandenbroucke-Grauls, C. M. Resistance to selective decontamination: the jury is still out. *Lancet Infect. Dis.* **13,** 282–283 (2013).
23. Schloss, P. D. Microbiology: An integrated view of the skin microbiome. *Nature* **514,** 44–45 (2014).
24. Laufer, A. S. *et al.* Microbial communities of the upper respiratory tract and otitis media in children. *MBio* **2,** e00245–e10 (2011).
25. Krismer, B. *et al.* Nutrient limitation governs Staphylococcus aureus metabolism and niche adaptation in the human nose. *PLoS Pathog.* **10,** e1003862 (2014).
26. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* **8,** 15–25 (2010).
27. Dobson, A., Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocin production: a probiotic trait? *Appl. Environ. Microbiol.* **78,** 1–6 (2012).
28. Kommineni, S. *et al.* Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. *Nature* **526,** 719–722 (2015).
29. Challinor, V. L. & Bode, H. B. Bioactive natural products from novel microbial sources. *Ann. NY Acad. Sci.* **1354,** 82–97 (2015).
30. Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158,** 1402–1414 (2014).
31. Heilbronner, S. *et al.* Genome sequence of *Staphylococcus lugdunensis* N920143 allows identification of putative colonization and virulence factors. *FEMS Microbiol. Lett.* **322,** 60–67 (2011).
32. Sidda, J. D. *et al.* Discovery of a family of gamma-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. *Chem. Sci.)* **5,** 86–89 (2014).
33. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13,** 509–523 (2015).
34. Sasaki, O., Igarashi, Y., Saito, N. & Furumai, T. Watasemycins A and B, new antibiotics produced by *Streptomyces* sp. TP-A0597. *J Antibiot. (Tokyo)* **55,** 249–255 (2002).
35. Miller, D. A., Luo, L., Hillson, N., Keating, T. A. & Walsh, C. T. Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chem. Biol.* **9,** 333–344 (2002).
36. Weber, T. *et al.* antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43,** W237–W243 (2015).
37. Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33,** 127–135 (2015).
38. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3,** 490–504 (2002).
39. Hobbs, J. K., Miller, K., O'Neill, A. J. & Chopra, I. Consequences of daptomycin-mediated membrane damage in *Staphylococcus aureus*. *J. Antimicrob. Chemother.* **62,** 1003–1008 (2008).
40. Tacconelli, E. & Kern, W. V. New antibiotics for skin and skin-structure infections. *Lancet Infect. Dis.* **14,** 659–661 (2014).
41. Zervos, M. J. *et al.* Epidemiology and outcomes of complicated skin and soft tissue infections in hospitalized patients. *J. Clin. Microbiol.* **50,** 238–245 (2012).
42. Weidenmaier, C., Goerke, C. & Wolz, C. *Staphylococcus aureus* determinants for nasal colonization. *Trends Microbiol.* **20,** 243–250 (2012).
43. Baur, S. *et al.* A nasal epithelial receptor for *Staphylococcus aureus* WTA governs adhesion to epithelial cells and modulates nasal colonization. *PLoS Pathog.* **10,** e1004089 (2014).
44. Andersen, P. S. *et al.* Influence of host genetics and environment on nasal carriage of *Staphylococcus aureus* in Danish middle-aged and elderly twins. *J. Infect. Dis.* **206,** 1178–1184 (2012).
45. Bieber, L. & Kahlmeter, G. *Staphylococcus lugdunensis* in several niches of the normal skin flora. *Clin. Microbiol. Infect.* **16,** 385–388 (2010).
46. Iwase, T. *et al. Staphylococcus epidermidis* Esp inhibits *Staphylococcus aureus* biofilm formation and nasal colonization. *Nature* **465,** 346–349 (2010).
47. Sanders, M. E. Impact of probiotics on colonizing microbiota of the gut. *J. Clin. Gastroenterol.* **45** (Suppl), S115–S119 (2011).
48. Becker, K., Heilmann, C. & Peters, G. Coagulase-negative staphylococci. *Clin. Microbiol. Rev.* **27,** 870–926 (2014).
49. Clemente, J. C., Ursell, L. K., Parfrey, L. W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **148,** 1258–1270 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank V. Winstel for technical assistance and A. Bobic, S. Heilbronner, W. Hoffmann, A. Jorge, D. Kretschmer, A. Kulik, M. Nega, E. Stegmann, V. Winstel, T. Weber, and W. Wohlleben for assistance and helpful discussions. Thanks to Bruker Daltonics for selected initial high-resolution mass spectrometry analysis and to T. Paululat for NMR experiments. This work was financed by German Research Council grants GRK1708 to S.G. and A.P.; TRR156, Schi510/8-1, and PE805/5-1 to B.S. and A.P.; TRR34 to C.W. and A.P; and SFB766 to C.W., H.B.-O., S.G., and A.P.; and by the German Center for Infection Research (DZIF) to C.W., A.P., B.K., M.W., and H.B.-O.

Author Contributions A.Z. isolated lugdunin, designed experiments and investigated the biological activities of lugdunin. M.C.K. purified lugdunin, designed experiments and determined the structure of lugdunin. D.J. identified IVK28 and performed transposon mutagenesis. A.B. designed and performed precursor incorporation studies. C.L. designed the human colonization study and analysed data. M.B., A.Z., C.S. and C.W. performed animal experiments, and M.W. and C.W. analysed data and performed statistical analysis. M.M. provided patient samples and supported MALDI-TOF analysis. N.A.S. and H.K. established total chemical synthesis of lugdunin. B.K. isolated lugdunin, analysed operon structure and performed bioinformatic analyses, A.Z., M.C.K., B.S., H.B.-O., S.G., A.P. and B.K. designed the study, analysed results, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.P. (andreas.peschel@uni-tuebingen.de).

Reviewer Information *Nature* thanks G. Challis, M. Gilmore and K. Lewis for their contribution to the peer review of this work.

METHODS

Data reporting. The animal experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. No statistical methods were used to predetermine sample size. For the human colonization study, only anonymous and randomized samples were obtained.

Strains and growth conditions. The *Staphylococcus* strains used in this study were *S. aureus* USA300 LAC, *S. aureus* USA300 NRS384, *S. aureus* Mu50, *S. aureus* RN4220, *S. aureus* SA113, *S. aureus* Newman, *S. aureus* PS187, *S. lugdunensis* IVK28, *S. lugdunensis* IVK28 Δ*lugD*, and *S. lugdunensis* IVK28-Xyl. Further strains used for MIC determination were *Enterococcus faecium* BK463, *E. faecalis* VRE366, *Listeria monocytogenes* ATCC19118, *Streptococcus pneumoniae* ATCC49619, *Pseudomonas aeruginosa* PAO1, and *Escherichia coli* DH5α. *E. coli* DC10B was used as the cloning host and *B. subtilis* 168 (*trpC2*) was used for precursor incorporation studies. In addition, a set of 60 *S. aureus* and 17 *S. lugdunensis* strains were isolated from diagnostic samples in the course of the colonization study described below.

Basic medium (BM: 1% soy peptone, 0.5% yeast extract, 0.5% NaCl, 0.1% glucose and 0.1% K_2HPO_4, pH 7.2) was used as the standard growth medium. MIC determinations and killing assays were performed in Mueller Hinton Broth (MHB; Roth, Karlsruhe, Germany). For the identification of *S. lugdunensis*, selective *S. lugdunensis* medium (SSL) was used as previously described[50]. When necessary, antibiotics were used at concentrations of 250 μg ml^{-1} for streptomycin, 10 μg ml^{-1} for chloramphenicol, 2.5 μg ml^{-1} for erythromycin, and 100 μg ml^{-1} for ampicillin.

Bioactivity test. The antimicrobial activity against *S. aureus* of *S. lugdunensis* IVK28 was identified by screening 90 nasal staphylococcal isolates for the capacity to inhibit growth of *S. aureus*. For this purpose, BM agar was inoculated 1:10,000 with an overnight culture of *S. aureus* USA300 LAC. The test strains were inoculated on the resulting bacterial lawn, and the plates were incubated for 24–48 h at 37 °C. To investigate the production of antimicrobial activity by IVK28 under iron-limiting conditions, BM agar was supplemented with 200 μM 2, 2'-bipyridine[25].

Transposon mutagenesis and elucidation of the lugdunin gene cluster. The temperature-sensitive plasmid pTV1ts, which contains the 5.3-kb transposon *Tn*917 (*erm*^{R}) from *E. faecalis*, was transferred into *S. lugdunensis* IVK28 by electroporation. Transposon mutants were screened for loss of antimicrobial activity against *S. aureus*. Chromosomal DNA was isolated by standard procedures from non-inhibitory clones, and the primers Tn917 up and Tn917 down (Extended Table 2) were used to directly sequence the flanking regions of the transposon insertion site. Sequence analysis was performed with DNASTAR Lasergene software (DNASTAR Inc., Madison, WI, USA). Bioinformatic analysis was performed by BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and antiSMASH 3.0 (ref. 36).

Generation of *S. lugdunensis* IVK28-Xyl. The flanking regions of *lugR* were amplified by PCR with the primer pairs SIPr1–up/SIPr1–down and SIPr2–up/SIPr2–down (Extended Data Table 2). The plasmid pBASE6-erm/lox 1, a derivative of pBASE6 (ref. 51), already containing an erythromycin resistance cassette in the singular SmaI site, was linearized with Acc65I. The identically digested SIPr1 PCR product, containing one natural Acc65I restriction site and one introduced by the primer, was ligated into pBASE6-erm/lox1. The resulting vector with the correctly oriented SIPr1 PCR product and the SIPr2 PCR product were ligated after digestion with EcoRV and BglII. The resulting pBASE6-erm/lox 1 construct with both flanking regions inserted was linearized with BssHII, treated with Klenow enzyme and digested with BglI. The required *xylR* fragment with the downstream-located *xylAB*-promoter was excised from pTX15 by HindIII restriction, treated with Klenow enzyme and subsequently digested with BamHI. The ligation of the *xylR* fragment into the appropriate vector generated pBASE6-erm/lox1-*xylR*, which was transferred into *E.coli* DC10B and subsequently into *S. aureus* PS187. The resulting plasmid pBASE6-erm/lox1-*xylR* was transduced into *S. lugdunensis* IVK28 via the bacteriophage Φ187 according to ref. 52. Homologous recombination for replacement of *lugR* by *erm/xylR* was performed as previously described[51], generating the xylose-inducible lugdunin producer strain *S. lugdunensis* IVK28-Xyl.

Production and purification of lugdunin. A fresh overnight culture of *S. lugdunensis* IVK28-Xyl was inoculated 1:1,000 in BM without glucose and was supplemented with 0.5% xylose. After incubation at 37 °C under continuous shaking (160 rpm) for 24 h, whole cultures were extracted with 1-butanol at a ratio of 5:1. The aqueous phase was discarded, and the organic phase was evaporated at 37 °C under reduced pressure and finally dissolved in methanol. The methanol extract was applied to a gel filtration column (Sephadex LH20, 1.6 × 80 cm, flow rate 1 ml min^{-1} methanol). The active fractions containing lugdunin were pooled, evaporated at 37 °C under reduced pressure and dissolved in dimethyl sulfoxide (DMSO). This solution was then subjected to a preparative reversed-phase HPLC column (Kromasil C18, 7 μm, 250 × 20 mm; Dr. Maisch, Ammerbuch, Germany) with an isocratic elution at 79% methanol in water for 20 min. The fractions containing lugdunin were baseline-separated from the remaining compounds. Methanol was evaporated at 37 °C under reduced pressure to yield lugdunin as a white powder.

Chemical synthesis of lugdunin. Total chemical synthesis was achieved by a Fmoc (9-fluorenylmethoxycarbonyl) strategy-based manual solid-phase peptide synthesis and was established on an H-Val-H NovaSyn TG resin (Novabiochem, Switzerland). Amino acids were coupled in a fourfold excess using HATU (1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate). Valine positions were coupled twice by PyOxim ([ethyl cyano(hydroxyimino)acetato-O^2]tri-1-pyrrolidinylphosphonium hexafluorophosphate) for the second coupling instead of HATU. Deprotection was performed in trifluoroacetic acid for 30 min. Peptides were cleaved from the resin with acetonitrile:water:trifluoroacetic acid (79.95:20:0.05) for 30 min. Lyophilisation yielded the crude product. Crude synthetic lugdunin product was purified by reversed-phase HPLC and compared with the natural product by electrospray ionization liquid chromatography high-resolution mass spectrometry, additional chiral-HPLC methods (column, Dr. Maisch Reprosil Chiral NR, Ammerbuch, Germany; elution with 80% premixed methanol in H_2O at 1.5 ml min^{-1} flow rate), bioactivity assay and advanced Marfey's analysis.

MIC assay, serum stability and spectrum of activity. *S. aureus* RN4220, *S. aureus* USA300 (LAC), *S. aureus* USA300 (NRS384), *S. aureus* SA113, *S. aureus* Mu50, *E. coli* DH5α and *P. aeruginosa* PAO1 were grown overnight in MHB. *E. faecalis* VRE366, *E. faecium* BK463, *S. pneumoniae*, and *L. monocytogenes* were grown in tryptic soy broth (Difco Laboratories, Augsburg, Germany). Serum stability of lugdunin was determined in 50% MHB with 50% human serum, which was obtained by standard Histopaque/Ficoll centrifugation of freshly isolated blood. All strains were incubated at 37 °C under continuous shaking. Early log-phase grown bacteria were adjusted in MHB to 1×10^6 c.f.u. per ml in microtiter plates, mixed with varying concentrations of the antibiotic and incubated at 37 °C for 24 h under continuous shaking. The OD_{600} of each well was measured with a microplate reader, and the lowest peptide concentrations, which displayed no bacterial growth, were defined as the MIC. The assays were performed in 96-well microtiter plates. MIC values for *B. subtilis* 168 to be used as points of reference in precursor incorporation studies were determined in Belitzky minimal medium[53], using a final inoculum of 5×10^5 c.f.u. per ml.

Killing assay. Fresh MHB was inoculated 1:10,000 with an overnight culture of *S. aureus* USA300 LAC and was incubated at 37 °C under continuous shaking (160 rpm) until bacteria were grown to 1×10^6 c.f.u. per ml. Then, 10 × MIC lugdunin was added. At the time points 0 h, 2 h, 4 h, 8 h, 24 h and 30 h, samples were taken and centrifuged. The pellet was resuspended in 1 × PBS and serially diluted. The dilutions were spotted on tryptic soy agar, and colony counts were determined after overnight incubation at 37 °C. To determine cell numbers $<10^2$ c.f.u. per ml, whole cultures were centrifuged and plated on tryptic soy agar.

Precursor incorporation studies. *B. subtilis* 168 (*trpC2*) is a widely used model organism for mode of action investigations and was also used for orienting studies on the mechanism of lugdunin[54–56]. To measure incorporation of radioactive precursors into acid-precipitable macromolecules, *B. subtilis* 168 (*trpC2*), exponentially growing in Belitzky minimal medium (OD_{600} of 0.04), was labelled with 0.02 MBq ml^{-1} of [methyl-^{3}H]thymidine (DNA synthesis), [5,6-^{3}H]uridine (RNA synthesis), [4,5-^{3}H]L-leucine (protein synthesis) or [1-^{3}H]glucosamine D-hydrochloride (cell-wall synthesis). After 5 min culture aliquots were treated with lugdunin at concentrations of 2, 1, 0.67 or 0.5 μg ml^{-1}, corresponding to 1/2×, 1/4×, 1/6× or 1/8× the MIC, respectively. Ciprofloxacin (2 μg ml^{-1}, 4 × MIC), rifampicin (1 μg ml^{-1}, 4 × MIC), chloramphenicol (16 μg ml^{-1}, 4 × MIC) or vancomycin (2 μg ml^{-1}, 4 × MIC) served as reference antibiotics for inhibition of DNA, RNA, protein or cell-wall biosynthesis, respectively. Samples (0.1 ml) were taken in regular time intervals and precipitated with 6% perchloric acid in a multiscreen filter plate (Millipore, 0.45 μm). After washing the precipitates with 0.15 ml of ethanol, plates were dried and radioactivity was determined with scintillation fluid (Ultima Gold, Perkin Elmer) in a 1450 MicroBeta TriLux counter (Wallac).

Cytotoxicity against eukaryotic cells. Human neutrophil granulocytes were freshly isolated from the blood of healthy volunteers by standard Histopaque/Ficoll centrifugation. Lysis of neutrophil granulocytes was monitored by the release of the enzyme lactate dehydrogenase (LDH), as described in ref. 57. Lugdunin was added at final concentrations of 50, 25 and 12.5 μg ml^{-1} in 0.5% DMSO to wells of a 96-well tissue culturing plate containing 1×10^6 neutrophil granulocytes per well in 200 μl RPMI-1640 medium (2 g l^{-1} $NaHCO_3$, 10% foetal calf serum, 1% L-glutamine and 1% penicillin-streptomycin, PAN Biotech) without phenol red. The plates were incubated at 37 °C and 5% CO_2 for 3 h and the lysis was determined with the LDH Cytotoxicity Detection Kit (Roche Applied Sciences, Mannheim, Germany).

Haemolytic activity was determined with human erythrocytes, freshly isolated from the blood of healthy volunteers by standard Histopaque/Ficoll centrifugation. Lugdunin was added at final concentrations of 50, 25 and 12.5 $\mu g\,ml^{-1}$ in 0.5% DMSO to 2% erythrocytes in $1\times$ PBS. The cells were incubated for 1 h and afterwards centrifuged for 10 min at 1,000 rpm. The supernatant was diluted 1:10 in $1\times$ PBS, and the absorbance was measured at a wavelength of 540 nm. As a positive control for neutrophil granulocyte or erythrocyte lysis, 2% Triton X-100 was added to the samples.

HL60 cells were cultured in RPMI-1640 medium. Their metabolic activity was monitored using the cell health indicator alamarBlue (Invitrogen) as described in ref. 58. Lugdunin was added at final concentrations of 50, 25 and 12.5 $\mu g\,ml^{-1}$ in 0.5% DMSO to wells of a 96-well tissue culturing plate containing 1×10^4 HL60 cells per well. The plates were incubated at 37 °C and 5% CO_2 for 24 h. As a positive control for high cytotoxicity, staurosporine was added to the samples. The HL60 cell line, which we have obtained from CNRS (Université Grenoble, France), and whose identity was confirmed by the provider, is not listed in the database as commonly misidentified. Mycoplasma assays were negative.

Resistance development study. MIC assays for the antibiotics used in this study were performed as described above. We determined $1\times$ MICs of 0.01 $\mu g\,ml^{-1}$ rifampicin and 1.5 $\mu g\,ml^{-1}$ lugdunin against *S. aureus* USA300. Fresh MHB was inoculated 1:10,000 with an overnight culture of *S. aureus* USA300 LAC and was incubated at 37 °C under continuous shaking. Cells were grown to early log phase, adjusted to 1×10^6 cells ml^{-1}, and dispensed into 96-well microtiter plates with 100 μl per well. Lugdunin and rifampicin were added at concentrations of $0.25\times$, $0.5\times$, $1\times$, $1.5\times$, $2\times$ and $4\times$ MIC. After 24 h incubation at 37 °C under continuous shaking, growth was determined with a microplate reader at an OD_{600}, and cells from the second highest concentration showing visible growth were used to inoculate the subsequent culture.

Statistical analyses. Statistical analysis was performed by using GraphPad Prism (GraphPad Software, Inc., La Jolla, USA; version 5.04). Statistically significant differences were calculated by appropriate statistical methods as indicated. For the human study, risk of nasal colonization with *S. aureus* in the presence or absence of *S. lugdunensis*, as well as the respective point estimates of the risk ratio and confidence intervals, were determined using Stata version 12.1 (Stat Corp., College Station, TX, USA). *P* values of ≤ 0.05 were considered significant.

Animal models and ethics statement. All animal experiments were conducted in strict accordance with the German regulations of the Gesellschaft für Versuchstierkunde/Society for Laboratory Animal Science (GV-SOLAS) and the European Health Law of the Federation of Laboratory Animal Science Associations (FELASA) in accordance with German laws after approval (protocol HT1/12 for mouse skin infection and T1/10 for cotton rat colonization) by the local authorities (Regierungspräsidium Tübingen). All, animal and human studies were carried out at the University Hospital Tübingen and conformed to institutional animal care and use policies. No randomization or blinding was necessary for the animal infection/colonization models, and no samples were excluded. Animal studies were performed with female C57BL/6 mice, 6–8 weeks old, or cotton rats of both genders, 8–10 weeks old, respectively. The human nasal colonization study was approved by the ethics committee of the medical faculty of the University Hospital Tübingen (project number 577/2015A).

Skin infection of C57BL/6 mice. A streptomycin-resistant *S. aureus* Newman strain was used to infect C57BL/6 mice epicutaneously by the tape-stripping technique from ref. 59. Tryptic soy broth with 500 $\mu g\,ml^{-1}$ streptomycin was inoculated 1:10,000 with a fresh overnight culture of the test strain and was incubated at 37 °C under continuous shaking until an OD_{600} of 0.5 was reached. Cells were centrifuged, washed twice with $1\times$ PBS, and adjusted to 1×10^8 cells per ml. The integrity of the shaved skin of the mice was affected by repeated (seven times) vigorous tape stripping to enable *S. aureus* Newman infection. An inoculum of 15 μl from the bacterial suspension was added to 7-mm filter paper discs, placed onto the prepared skin with two discs per animal, and covered with Finn chambers on Scanpor tape (Smart Practise, Phoenix, AZ, USA). Finn chamber fixation occurred via Fixomull stretch plasters (BSN medical GmbH, Hamburg, Germany). After incubation for 24 h, the Finn chambers were removed and 1.5 μg of lugdunin per colonized area was applied, followed by a second and third treatment with the same amount of lugdunin after 30 h and 42 h. Six hours after the final application, mice were euthanized, the skin was large-scale detached and 4-mm punches of the originally colonized areas were vortexed in $1\times$ PBS for 30 s to remove the attached bacteria from the skin (wash fraction). The skin was dissected with a scalpel to expose bacteria from deeper areas of the skin (tissue fraction), which was homogenized by vortexing in $1\times$ PBS for 30 s. c.f.u. of both fractions were determined by serial dilutions in $1\times$ PBS, which were spotted onto tryptic soy agar, supplemented with streptomycin, for a streptomycin-resistant *S. aureus* Newman-specific selection. The plates were incubated overnight at 37 °C.

Generation of *S. lugdunensis* Δ*lugD* and complementation. For the construction of a marker-less knockout strain, 1-kb flanking regions of *lugD* were amplified by PCR with the primer pairs *lugD* upstream-SacI/*lugD* upstream-Acc65I and *lugD* downstream-Acc65I/*lugD* downstream-BglII (Extended Data Table 2). The fragments were digested according to their introduced restriction sites and were ligated into the plasmid pBASE6 generating pBASE6-Δ*lugD*, which was transferred to *E. coli* DC10B. The correct plasmid was transferred to *S. aureus* PS187 by electroporation, which was then infected with the bacteriophage Φ187 for the transduction of pBASE6-Δ*lugD* into *S. lugdunensis* IVK28 wild type. The knockout was generated by homologous recombination of the flanking regions into the genome, and deletion of *lugD* was confirmed by PCR. For the complementation of the mutant, *lugD* was amplified by the primer pair *lugD* complementation forward-PstI/*lugD* complementation reverse-Acc65I (Extended Data Table 2), digested with the appropriate restriction enzymes and ligated into identically digested pRB474. The resulting pRB474-*lugD* was transduced into *S. lugdunensis* IVK28 Δ*lugD*, as described for the knockout mutant.

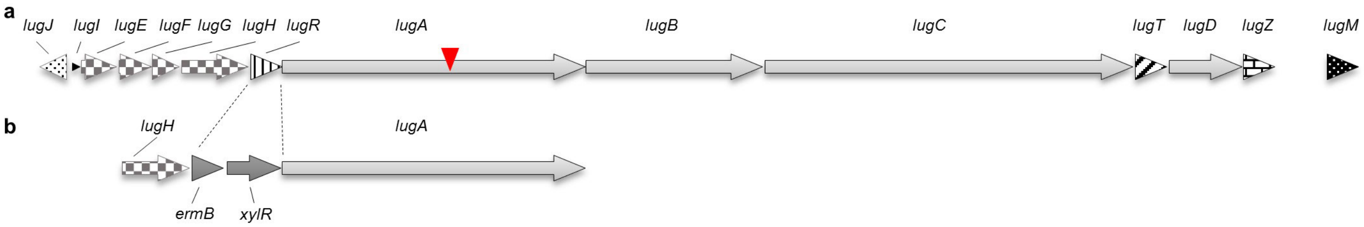
Competition assay. *S. lugdunensis* IVK28 wild type, *S. lugdunensis* IVK28 Δ*lugD, S. lugdunensis* IVK28 Δ*lugD*::pRB474-lugD, and a streptomycin-resistant *S. aureus* Newman were grown in BM overnight at 37 °C under continuous shaking. These strains were then adjusted to 1×10^9 c.f.u. per ml in $1\times$ PBS and diluted 1:10. For the starting condition of 90% *S. aureus*, equal volumes of 1×10^9 *S. aureus* c.f.u. per ml and 1×10^8 *S. lugdunensis* c.f.u. per ml were mixed. Co-cultures with only 10% *S. aureus* were also performed, and 20 μl of these mixtures were spotted in triplicate on BM agar and incubated at 37 °C. Samples were taken at 0 h, 24 h, 48 h, and 72 h by scraping cells from the agar plates and suspending them in $1\times$ PBS. Serial dilutions of these samples were plated on BM and BM containing streptomycin for selection of *S. aureus*. After overnight incubation at 37 °C, colony counts were determined, and the bacterial ratios of *S. aureus* and *S. lugdunensis* were calculated.

Co-colonization of cotton rat noses. For the colonization of cotton rat noses, spontaneous streptomycin-resistant mutants of *S. lugdunensis* IVK28 wild type and *S. lugdunensis* IVK28 Δ*lugD* were selected on BM agar plates containing 250 $\mu g\,ml^{-1}$ streptomycin. Co-colonization was conducted with a streptomycin-resistant *S. aureus* Newman. The cotton rat model was described previously[43]. As the capacity of *S. lugdunensis* to colonize cotton rat nares has not been studied before, we determined the inoculum required for stable colonization by IVK28 wild type and its mutant Δ*lugD* over 5 days. Our previous studies have shown that for *S. aureus* an inoculum of 10^7 bacteria per nose results in a stable colonization of about 10^3 c.f.u. per nose (Extended Data Fig. 7a). To achieve a comparable colonization level with *S. lugdunensis*, an inoculum of 10^8 bacteria per nose was required, and there was no detectable difference in colonization efficiency between wild type and Δ*lugD* (Extended Data Fig. 7b, c). Therefore, co-colonization experiments in cotton rat noses were performed with tenfold more *S. lugdunensis* than *S. aureus* to obtain a 1:1 colonization ratio.

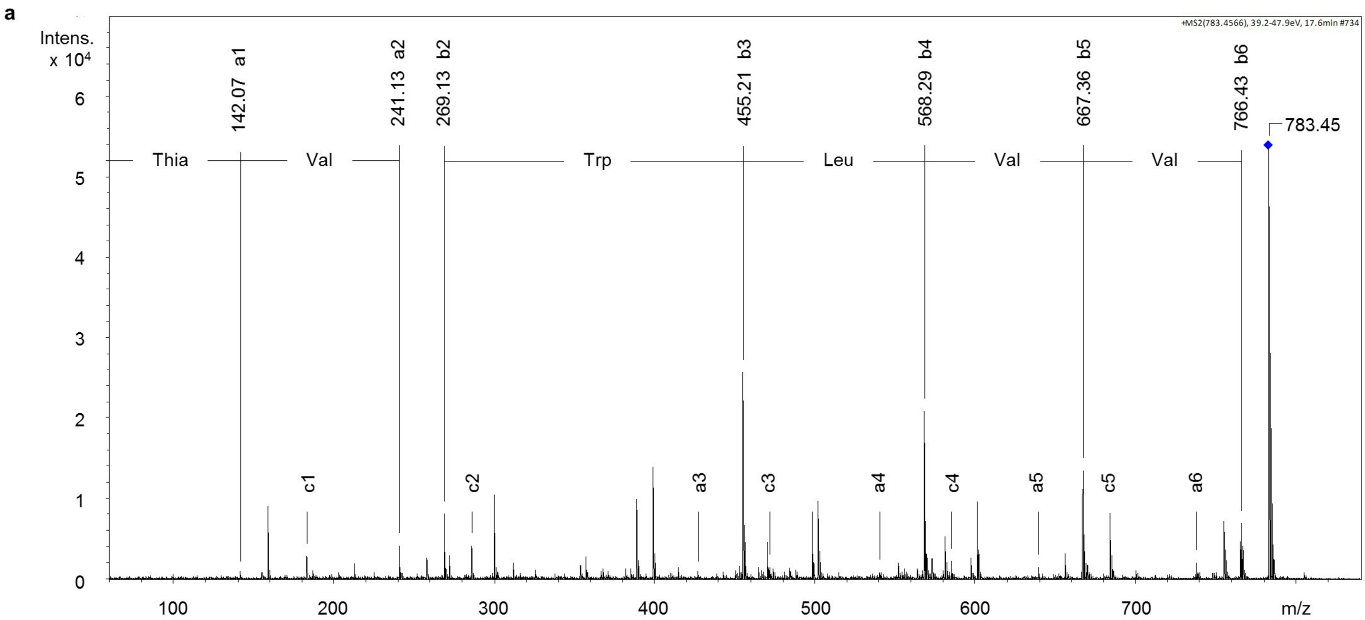
Cotton rats were anaesthetized and instilled intranasally with mixtures of either 1×10^8 *S. lugdunensis* wild type and 1×10^7 *S. aureus* Newman or 1×10^8 *S. lugdunensis* Δ*lugD* and 1×10^7 *S. aureus* Newman. Five days after bacterial instillation, the animals were euthanized, and noses were surgically removed. The noses were heavily vortexed in 1 ml of $1\times$ PBS for 30 s. Dilutions of the samples in PBS were plated on SSL agar containing 250 $\mu g\,ml^{-1}$ streptomycin to select for the used strains and to separate *S. aureus* (yellow) and *S. lugdunensis* (purple) by colour. The plates were incubated for 2 days under anaerobic conditions (anaerobic jar with Anaerocult A, MerckKGaA), for the specific detection of ornithine decarboxylase activity. *S. aureus* Newman c.f.u. were determined afterwards. All animals received drinking water with 2.5 $mg\,ml^{-1}$ streptomycin continuously, starting 3 days before the experiment, to reduce the natural nasal flora.

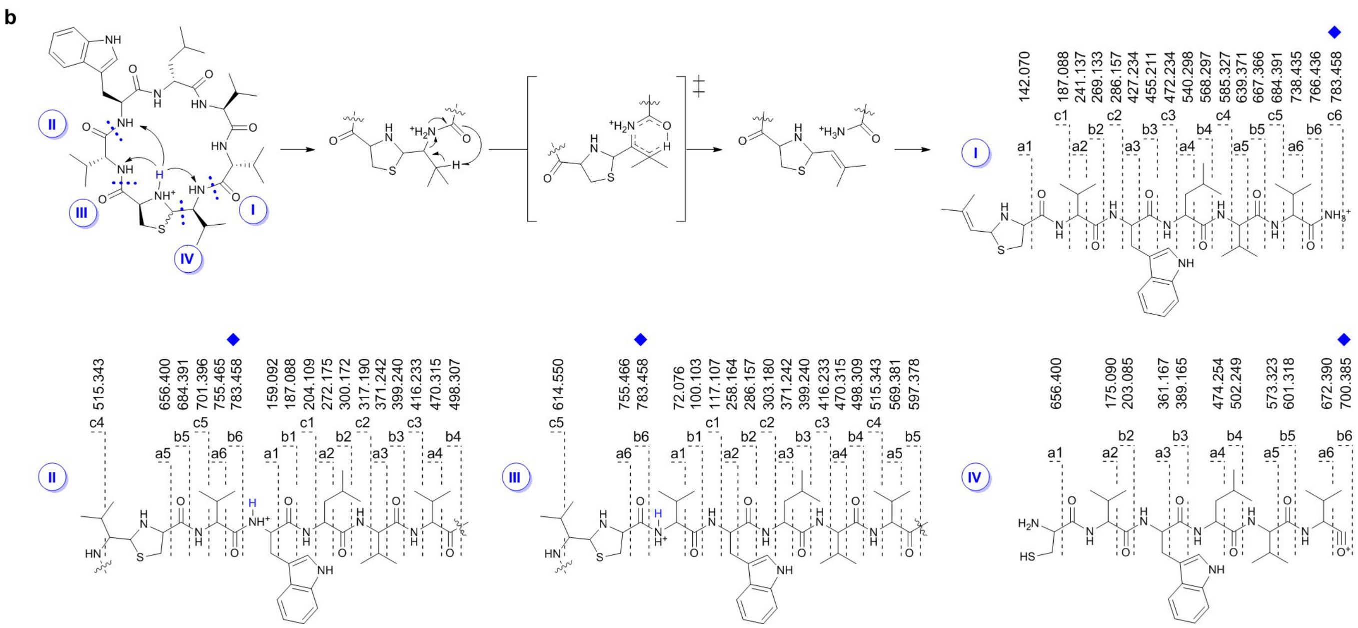
Human colonization study. A total of 187 nasal swab samples from hospitalized patients were received from the diagnostics laboratory of the Institute of Medical Microbiology and Hygiene (University Hospital Tübingen, Germany). Dilutions from each sample were plated on blood agar and SSL agar for a phenotypic identification of *S. aureus* and *S. lugdunensis*. Identity was confirmed by coagulase test and matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (mass spectrometer: AXIMA Assurance, Shimadzu Europa GmbH, Duisburg, database: SARAMIS with 23.980 spectra and 3.380 superspectra, BioMérieux, Nuertingen).

50. Ho, P. L. *et al.* Novel selective medium for isolation of *Staphylococcus lugdunensis* from wound specimens. *J. Clin. Microbiol.* **52,** 2633–2636 (2014).
51. Geiger, T. *et al.* The stringent response of *Staphylococcus aureus* and its impact on survival after phagocytosis through the induction of intracellular PSMs expression. *PLoS Pathog.* **8,** e1003016 (2012).
52. Winstel, V., Kühner, P., Krismer, B., Peschel, A. & Rohde, H. Transfer of plasmid DNA to clinical coagulase-negative staphylococcal pathogens by using a unique bacteriophage. *Appl. Environ. Microbiol.* **81,** 2481–2488 (2015).
53. Wenzel, M. *et al.* Small cationic antimicrobial peptides delocalize peripheral membrane proteins. *Proc. Natl Acad. Sci. USA* **111,** E1409–E1418 (2014).
54. Bandow, J. E., Brötz, H., Leichert, L. I., Labischinski, H. & Hecker, M. Proteomic approach to understanding antibiotic action. *Antimicrob. Agents Chemother.* **47,** 948–955 (2003).
55. Hutter, B. *et al.* Prediction of mechanisms of action of antibacterial compounds by gene expression profiling. *Antimicrob. Agents Chemother.* **48,** 2838–2844 (2004).
56. Sass, P. *et al.* Antibiotic acyldepsipeptides activate ClpP peptidase to degrade the cell division protein FtsZ. *Proc. Natl Acad. Sci. USA* **108,** 17474–17479 (2011).
57. Wang, R. *et al.* Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nat. Med.* **13,** 1510–1514 (2007).
58. Bara, R. *et al.* Atropisomeric dihydroanthracenones as inhibitors of multiresistant *Staphylococcus aureus*. *J. Med. Chem.* **56,** 3257–3272 (2013).
59. Wanke, I. *et al.* *Staphylococcus aureus* skin colonization is promoted by barrier disruption and leads to local inflammation. *Exp. Dermatol.* **22,** 153–155 (2013).

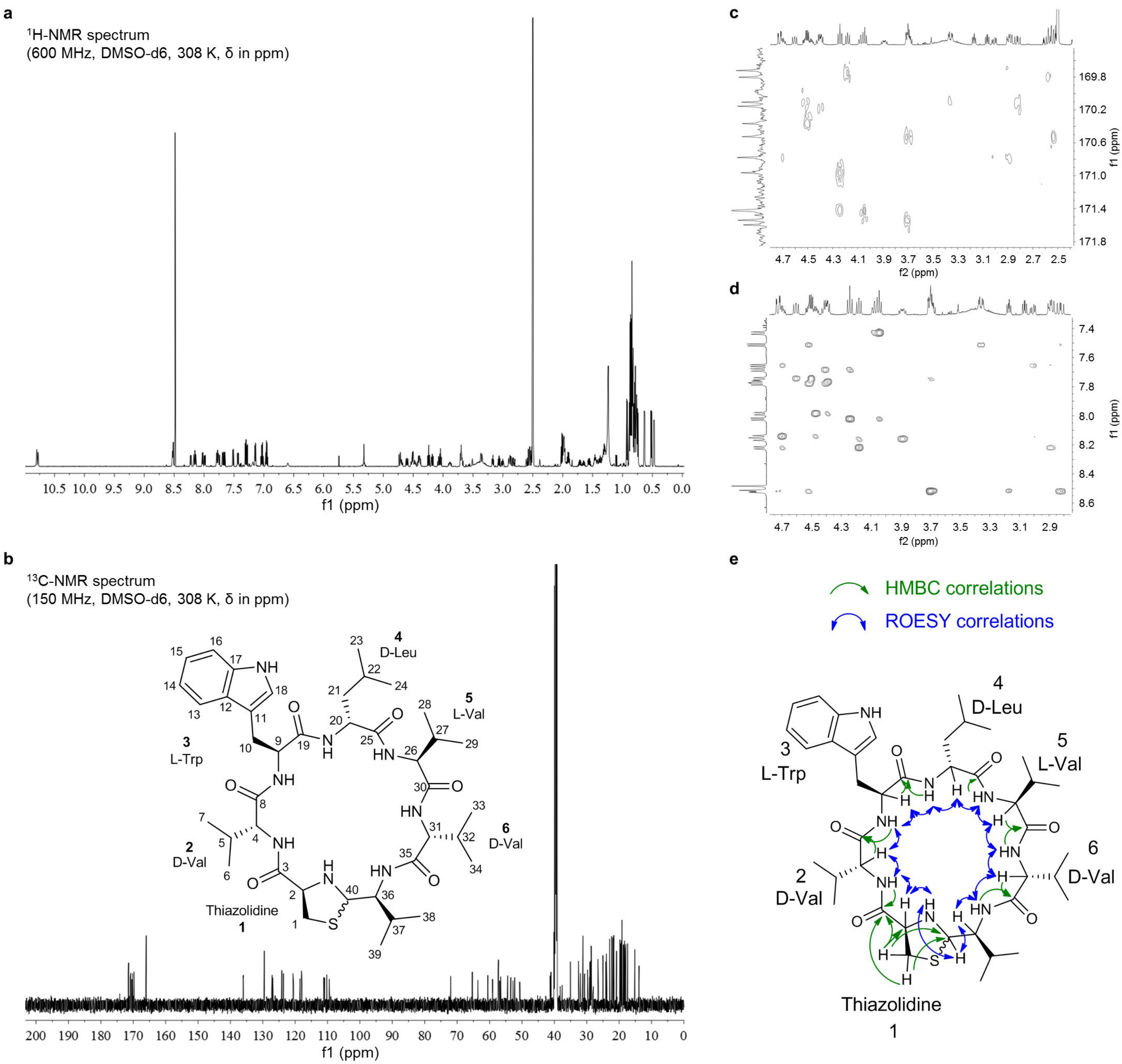


Extended Data Figure 1 | Gene cluster of lugdunin and generation of *S. lugdunensis* IVK28-Xyl. a, The lugdunin genes are located on a 30-kbp operon. *lugA–D* encode the four NRPS, which are preceded by the putative regulator gene *lugR*. Encoded upstream are putative ABC transporter genes (*lugE–H*) and genes for proteins with no described function (*lugI* and *lugJ*). A type-II thioesterase, that functions as a repair enzyme for stalled PCP domains, is encoded between *lugC* and *lugD*. The 4'-phosphopantetheinyl transferase LugZ converts inactive PCPs (apo-PCP) into the active holo-form by attachment of the 4'-phosphopantetheine cofactor. Encoded downstream is a putative monooxygenase (*lugM*). The transposon insertion site of Tn917, generating the lugdunin-deficient mutant IVK28 M1, is indicated by a red arrow. **b**, The xylose-inducible lugdunin producer strain *S. lugdunensis* IVK28-xyl was generated by replacement of the regulator gene *lugR* by the *xylAB* promoter along with the *xylR* gene encoding a xylose-sensitive repressor. The erythromycin resistance cassette *ermB* was integrated for selection purposes.

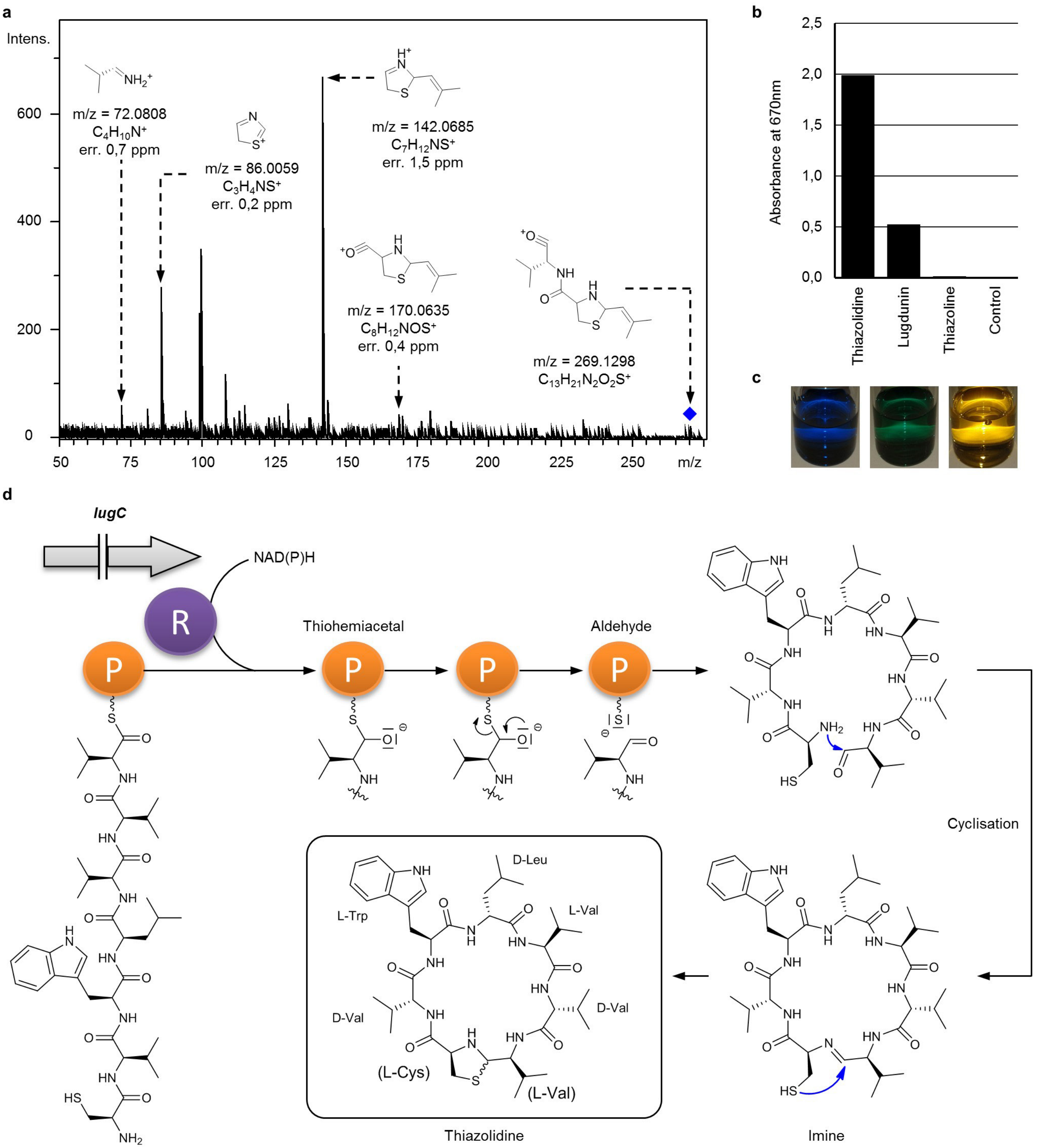




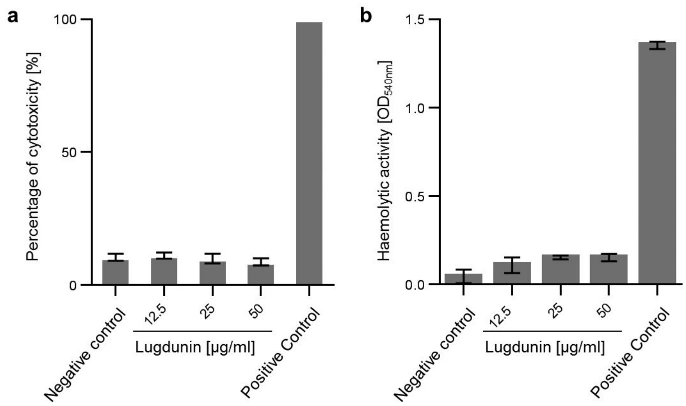
Extended Data Figure 2 | Structure elucidation by multistage tandem and electrospray ionization high-resolution mass spectrometry. **a**, A single-stage high-resolution MS/MS experiment revealed a superposition of fragment ions typically found for cyclic peptides. Quasi molecular ion selected for fragmentation is marked with a blue rhombus ($m/z = 783.45$). For sequence annotation, fragment ions were searched for b-ions of high intensity. One fragmentation route is annotated exemplarily by highlighting respective a-, b- and c-ion series signals. Thia, thiazoldine. **b**, Data generated by multistage tandem mass spectrometry show that lugdunin mainly fragmented along four routes (blue digits; (I), (II), (III), (IV)). Initial protonation occurs at the secondary amine of thiazolidine (blue H atom). Proton transfer to nearby peptide bonds initiated ring cleavage with subsequent fragmentation. Initial loss of ammonia for fragmentation route (I) can be explained by a precedent six-membered transition state. Intensities are in arbitrary units. Blue rhombuses label the position of initial ring cleavage. Fragmentation route molecules are shown in linearized form.



Extended Data Figure 3 | NMR spectra of the natural product lugdunin. The two diastereomeric and interconvertible forms of lugdunin (see imine intermediate, Extended Data Fig. 4d) show distinctive sets of NMR signals with consistent patterns, which were assigned by two-dimensional NMR methods (COSY, HMBC, HSQC-DEPT, ROESY and TOCSY) and corroborated the thiazolidine heterocycle as a special feature of lugdunin. Chemical shifts (δ) are shown in p.p.m. **a**, ^{1}H NMR spectrum (600 MHz) of lugdunin in DMSO-d6 at 308 K. **b**, ^{13}C NMR spectrum (150 MHz) of lugdunin in DMSO-d6 at 308 K. Atom numbering refers to full spectral assignment shown in Extended Data Table 1. **c**, Expansion of HMBC spectrum (heteronuclear multiple bond correlation) shows distinct correlations between the amino acid α-proton and the carbonyl C-atom of the respective amino acid. **d**, Expansion of ROESY spectrum (rotation frame nuclear Overhauser effect spectroscopy) shows short-ranged correlations through space between α-protons and amino acid amide protons. **e**, Taken together, HMBC (green arrows) and ROESY (double headed blue arrows) correlations allowed for a full sequential walk along the peptide backbone, which readily confirmed the amino acid sequence of both lugdunin diastereomers (the sequential walk is exemplarily shown for one diastereomer).



Extended Data Figure 4 | The thiazolidine moiety of lugdunin and its formation by peptide cyclisation. a, Blue rhombus marks the b2-ion from fragmentation route I (Extended Data Fig. 2). The ion was selected from an in-source collision-induced decay (isCID) experiment and was further fragmented in order to find thiazolidine-specific fragment ions. High-resolution MS data are shown with annotated sum formulas and respective deviations from calculated masses (errors in p.p.m.). Shown fragment ions reveal the b1-ion at $m/z = 170.0635$ Da. Intensity is shown in arbitrary units. **b**, Photometric detection of thiazolidines at 670 nm. A positive colour reaction with 3-methyl-2-benzothiazolone hydrazone hydrochloride (Sawicki reagent) gives rise to detectable absorbance at 670 nm. Thiazolidine-4-carboxylic acid (thiazolidine) was chosen as positive control, whereas 2-methyl-2-thiazoline (thiazoline) and DMSO acted as negative controls. **c**, Observable colours for the thiazolidine detection reaction. Thiazolidine-4-carboxlic acid (blue), lugdunin (green) and negative controls (yellow). **d**, The terminal reductase of LugC is proposed to initiate cleavage of the thioester-bound peptide chain with the aid of an NAD(P)H cofactor. The mature heptapeptide is liberated reductively from the NRPS multienzyme complex and cyclises via the N-terminal amine (L-Cys) and C-terminal aldehyde (L-Val) to form a macrocyclic imine/Schiff base. Subsequent nucleophilic attack of the cysteine thiol group generates the five-membered thiazolidine heterocycle. Nucleophilic attack by the L-cysteine sulfhydryl group may occur either at the *re* or *si* face of the imine thus leading to a diastereomeric mixture of two structural populations (depicted with wavy bond). *LugC* is shown as truncated gene, as indicated by the two lines (||).

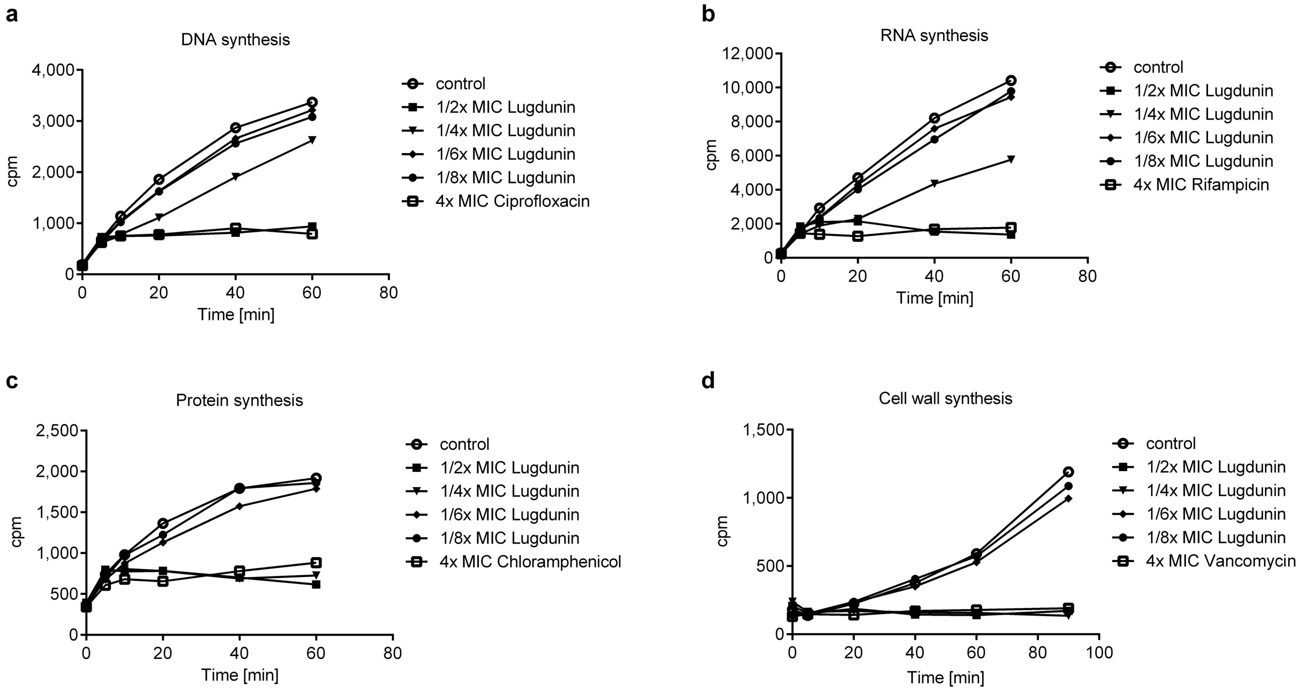


c

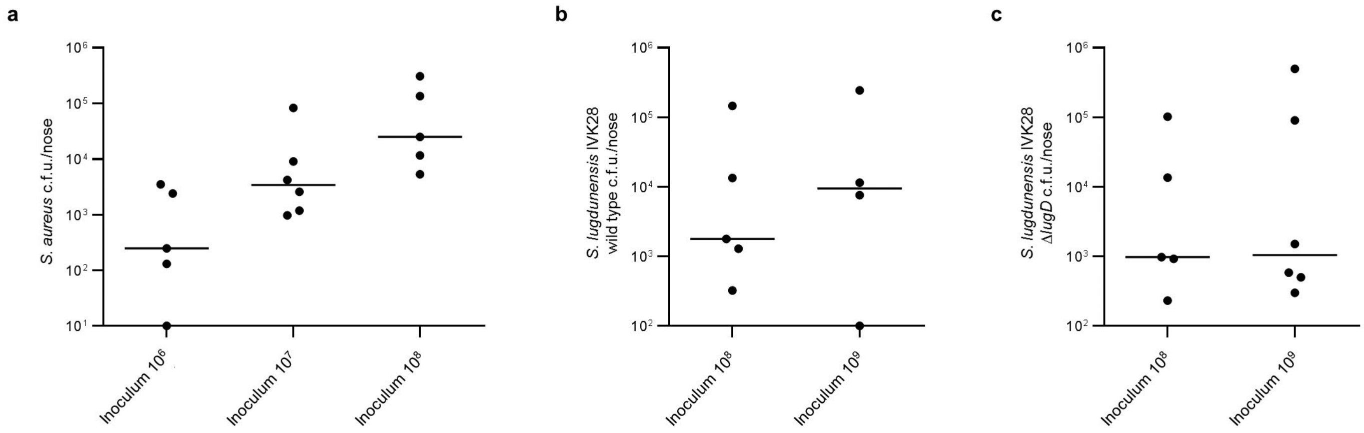
	HL60 - promyelocytic leukemia cells
Compound	IC_{50} [µg/ml]
Lugdunin	> 50
Staurosporine	2

IC_{50}, 50 % inhibitory concentration

Extended Data Figure 5 | Lugdunin activity against eukaryotic cells. **a**, **b**, Human neutrophil granulocytes (**a**) or erythrocytes (**b**) were incubated with high concentrations of lugdunin. Their lysis was monitored by the release of the enzyme lactate dehydrogenase (**a**) or haemoglobin (**b**), respectively. Cells without lugdunin were used as negative control. Incubation of cells in 2% Triton X-100 was used as positive control for high lysis. The data represent three independent experiments ± s.d. **c**, Promyelocytic leukaemia (HL60) cells were incubated with high concentrations of lugdunin. In order to determine the 50% inhibitory concentration of lugdunin, the metabolic activity of HL60 cells was measured by the conversion of resazurin into the highly fluorescent resorufin. The apoptosis-inducing staurosporine was used as positive control. IC_{50} values were calculated from the means of three independent experiments.



Extended Data Figure 6 | Time course of incorporation of tritium-labelled metabolic precursors into lugdunin-treated *B. subtilis*. *B. subtilis* 168 (trpC2) is a widely used model organism for mode of action investigations and was also used for orienting studies on the mechanism of lugdunin. Susceptibility of *B. subtilis* and *S. aureus* to lugdunin is similar (Table 1), with an MIC for *B. subtilis* of $4\,\mu g\,ml^{-1}$ in the Belitzky minimal medium used in this assay. **a**–**d**, Incorporation of thymidine into DNA (**a**), uridine into RNA (**b**), leucine into protein (**c**) or *N*-acetylglucosamine into peptidoglycan (**d**) ceased within minutes during lugdunin treatment. The experiment was repeated on three days with three independent bacterial cultures. One representative experiment is shown. At a concentration of half the MIC (1/2 MIC), incorporation of precursors into all pathways ceased reproducibly. At 1/8 MIC, incorporation continued repeatedly in parallel with the untreated control. At 1/4 MIC, the depicted experiment shows protein and peptidoglycan syntheses slightly more impaired than DNA and RNA syntheses, whereas the opposite occurred in another experiment. In summary, all four metabolic pathways seem to be equally impaired by lugdunin.



Extended Data Figure 7 | Nasal colonization rates by *S. aureus* and *S. lugdunensis* in cotton rats. a–c, Different inocula of *S. aureus* Newman (**a**), *S. lugdunensis* IVK 28 wild type (**b**) and *S. lugdunensis* IVK 28 Δ*lugD* (**c**) were instilled intranasally to determine their efficiency to colonize the noses of cotton rats (5 or 6 animals per group). c.f.u. of each strain were determined per nose after 5 days and plotted as individual dots. Lines represent the median of each group.

Extended Data Table 1 | NMR spectral assignment of lugdunin diastereomers

	Diastereomer I		Diastereomer II		
C-Atom	**δc**	**δH (mult., *J* in Hz)**	**δc**	**δH (mult., *J* in Hz)**	**C-Atom**
		1 L-Cys-Thiazolidine			
2-NH	-	2.88 (1H, m)	-	n.d.	2-NH
1	38.2	2.58 (1H, m) 3.06 (1H, dd, 9.7, 6.4)	37.5	2.54 (1H, m) 3.17 (1H, dd, 9.2, 6.0)	1
2	63.3	3.89 (1H, m)	65.3	3.69 (1H, m)	2
3	169.8	-	171.5	-	3
		2 D-Val			
4-NH	-	8.16 (1H, d, 9.8)	-	8.52 (1H, d, 8.6)	4-NH
4	57.2	4.18 (1H, dd, 10.0, 8.4)	60.5	3.70 (1H, m)	4
5		1.65 (1H, m)	28.5	1.71 (1H, m)	5
6, 7	18.6, 18.0	0.47 (3H, d, 6.6), 0.63 (3H, d, 6.6)	18.5, 19.3	0.52 (3H, d, 6.8), 0.84 (3H, m)*	6, 7
8	169.7	-	170.5	-	8
		3 L-Trp			
9-NH	-	8.22 (1H, d, 8.5)	-	8.52 (1H, d, 7.4)	9-NH
9	52.7	4.70 (1H, ddd, m)	53.4	4.52 (1H, m)	9
10	28.0	2.89 (1H, m) 3.01 (1H, dd, 14.0, 5.5)	26.4	2.82 (1H, dd, 14.7, 4), 3.36 (1H, dd, 14.7, 4)	10
11	109.5	-	110.2	-	11
12	127.1	-	126.8	-	12
13	118.2	7.66 (1H, d, 7.9)	117.9	7.51 (1H, d, 7.9)	13
14	117.9	6.95 (1H, ddd, 7.9, 7.8, 1.1)	118.1	6.96 (1H, ddd, 7.9, 7.9, 1.0)	14
15	120.5	7.02 (1H, ddd, 7.9, 7.8, 1.1)	120.7	7.04 (1H, ddd, 8.0, 7.9, 1.0)	15
16	110.9	7.28 (1H, dd, 7.9, 1.0)	111.1	7.30 (1H, dd, 8.0, 1.0)	16
17	136.0	-	136.1	-	17
18-NH	-	10.77 (1H, d, 2.0)	-	10.80 (1H, d. 2.0)	18-NH
18	124.1	7.15 (1H, d, 2.0)	123.6	7.15 (1H, d, 2.0)	18
19	170.8	-	170.1	-	19
		4 D-Leu			
20-NH	-	8.14 (1H, d, 8.7)		7.78 (1H, d, 9.4)	20-NH
20	50.7	4.46 (1H, ddm)	52.2	4.41 (1H, m)	20
21	41.3	1.32 (1H, m), 1.40 (1H, m)	41.1	1.30 (1H, m), 1.37 (1H, m)	21
22	24.1	1.28 (1H, m)	23.9	1.24 (1H, m)	22
23, 24	19.1, 19.3	0.84 (3H, m), 0.87 (3H, m)	20.8*, 21.6*	0.79 (3H, m), 0.77 (3H, m)	23, 24
25	171.6	-	171.4	-	25
		5 L-Val			
26-NH	-	7.99 (1H, d, 9.2)	-	7.69 (1H, d, 9.3)	26-NH
26	56.9	4.39 (1H, m)	57.4	4.24 (1H, dd, 9.2, 9.2)	26
27	31.2	1.96 (1H, m)	29.7	1.89 (1H, m)	27
28, 29	19.1, 17.9	0.80 (3H, m), 0.83 (3H, m)	19.1, 19.3	0.85 (3H, m)*, 0.84 (3H, m)*	28, 29
30	170.2	-	171.0	-	30
		6 D-Val			
31-NH	-	7.77 (1H, d, 8.7)	-	8.02 (1H, d, 7.8)	31-NH
31	56.5	4.50 (1H, dd, 8.8, 4.6)	59.1	4.04 (1H, dd, 7.8, 7.8)	31
32	31.9	1.97 (1H, m)	29.0	1.98 (1H, m)	32
33, 34	17.4, 19.1	0.83 (3H, d, 6.5), 0.88 (3H, d, 6.5)	18.3, 19.1	0.88 (3H, m), 0.84 (3H, m)	33, 34
35	170.4	-	171.4	-	35
		7 L-Val-Thiazolidine			
36-NH	-	7.74 (1H, d, 9.2)	-	7.43 (1H, d, 9.3)	36-NH
36	57.2	3.70 (1H, m)	54.4	4.04 (1H, dm, 9.3)	36
37	29.2	1.99 (1H, m)	32.5	1.55 (1H, m)	37
38, 39	15.2,22.3	0.75 (3H, d, 6.9), 0.79 (3H, d, 6.9)	19.8,19.7	0.86 (3H, d, 6.7), 0.92 (3H, d, 6.7)	38, 39
40	72.0	4.60 (1H, dd, 10.6, 8.8)	72.1	4.73 (1H, dd, 13.1, 2.0)	40

Chemical shifts (*δ*) are given in p.p.m. Mult., multiplicity. Coupling constants (*J*) are given in Hertz (Hz). Asterisks (*) mark strong signal overlap.

Extended Data Table 2 | Primers used in this study

Primer	Primer sequence 5'-3'	Gene annotation in *S. lugdunensis* N920143	Application for
Tn917 up	ATAGGCCTTGAAACATTGGTTTAGTGGG	---	Sequencing of transposon insertion site
Tn917 down	CCCATAGATAAGAAATACACCTGCAATAACC	---	Sequencing of transposon insertion site
SlPr1-up	TACGGTACCCGCTTAACAAGATGACTAGC	SLUG_RS03935	Replacement of *lugR* by *xylAP* promoter and *xylR* regulator gene
SlPr1-down	TCTTTATGGTACCTATTACATCTCTAAAG	SLUG_ RS03935	Replacement of *lugR* by *xylAP* promoter and *xylR* regulator gene
SlPr2-up	ATTTGTATTGATATCATAAAAAATGTCCG	SLUG_ RS03935	Replacement of *lugR* by *xylAP* promoter and *xylR* regulator gene
SlPr2-down	GTTAGATCTAAAGGAGGTCAATCAGATGG	SLUG_ RS03935	Replacement of *lugR* by *xylAP* promoter and *xylR* regulator gene
lugD upstream-SacI	TAGGAGCTCGCTTAATGAATTC	SLUG_RS03965	*S. lugdunensis* Δ*lugD*
lugD upstream-Acc65I	ATAGGTACCCTCCTTCTAGCTAAGC	SLUG_RS03965	*S. lugdunensis* Δ*lugD*
lugD downstream-Acc65I	AGTGGTACCCTCTATTAACTGAAAGG	SLUG_RS03965	*S. lugdunensis* Δ*lugD*
lugD downstream-BglII	ATTAGATCTGAAGTTAAGCATCCGTC	SLUG_RS03965	*S. lugdunensis* Δ*lugD*
lugD comp. forw-PstI	ATACTGCAGGCTTAGCTAGAAGGAGAG	SLUG_RS03965	Complementation of *S. lugdunensis* Δ*lugD*
lugD comp. rev-Acc65I	AATGGTACCCATCAGCATTATAGTT	SLUG_RS03965	Complementation of *S. lugdunensis* Δ*lugD*

ARTICLE

doi:10.1038/nature18934

Structural basis of Smoothened regulation by its extracellular domains

Eamon F. X. Byrne[1]*, Ria Sircar[2]*, Paul S. Miller[1], George Hedger[3], Giovanni Luchetti[2], Sigrid Nachtergaele[2], Mark D. Tully[4], Laurel Mydock-McGrane[5], Douglas F. Covey[5], Robert P. Rambo[4], Mark S. P. Sansom[3], Simon Newstead[3]*, Rajat Rohatgi[2] & Christian Siebold[1]

Developmental signals of the Hedgehog (Hh) and Wnt families are transduced across the membrane by Frizzled-class G-protein-coupled receptors (GPCRs) composed of both a heptahelical transmembrane domain (TMD) and an extracellular cysteine-rich domain (CRD). How the large extracellular domains of GPCRs regulate signalling by the TMD is unknown. We present crystal structures of the Hh signal transducer and oncoprotein Smoothened, a GPCR that contains two distinct ligand-binding sites: one in its TMD and one in the CRD. The CRD is stacked atop the TMD, separated by an intervening wedge-like linker domain. Structure-guided mutations show that the interface between the CRD, linker domain and TMD stabilizes the inactive state of Smoothened. Unexpectedly, we find a cholesterol molecule bound to Smoothened in the CRD binding site. Mutations predicted to prevent cholesterol binding impair the ability of Smoothened to transmit native Hh signals. Binding of a clinically used antagonist, vismodegib, to the TMD induces a conformational change that is propagated to the CRD, resulting in loss of cholesterol from the CRD–linker domain–TMD interface. Our results clarify the structural mechanism by which the activity of a GPCR is controlled by ligand-regulated interactions between its extracellular and transmembrane domains.

The extracellular region of Smoothened (SMO) is composed of an N-terminal CRD followed by a small linker domain, which then connects to the TMD and a C-terminal intracellular domain (ICD; Fig. 1a). Studies using small-molecule agonists and antagonists of SMO have defined two separable ligand-binding sites, one in the TMD and one in the CRD[1]. The TMD binding site binds the plant-derived inhibitor cyclopamine[2,3], the synthetic agonist SAG[4,5], and the anti-cancer drug vismodegib[6], which is used clinically to treat advanced basal cell cancer. Side-chain oxysterols such as 20(S)-hydroxycholesterol (20(S)-OHC) represent a distinct class of SMO ligands[7–9] that activate signalling by engaging a hydrophobic groove on the surface of the SMO CRD[10–12]. The native morphogen Sonic Hedgehog (SHH) functions by binding and inactivating Patched 1 (PTCH1), the major receptor for Hh ligands, which suppresses SMO activity[13]. Despite the discovery of numerous exogenous SMO ligands, no *bona fide* endogenous SMO ligand that regulates Hh signalling has been identified. Structure-guided mutations that disrupt 20(S)-OHC binding to the CRD groove or sterol-based inhibitors that occlude this groove impair signalling by SHH[10,11]. By contrast, several mutations in the TMD site that blocked the binding and activity of synthetic ligands failed to have any effect on the basal or SHH-stimulated activity of SMO[12,14]. These data suggest that an endogenous SMO ligand that can regulate Hh signalling engages the CRD groove on SMO.

Crystal structures of the isolated SMO linker domain–TMD in complex with both agonist and antagonist ligands[15–17] have shown that the GPCR heptahelical scaffold is conserved and provided a detailed view of a small-molecule binding pocket, but did not show the conformational changes typically associated with GPCR signalling[18,19]. In addition, two unliganded structures of the isolated SMO CRD have been solved[10,20]. However, structural insights into how the extracellular domains and TMD interact to regulate signalling in SMO (or in any other GPCR) are lacking.

Overall structure of SMO

We determined the crystal structure of human SMO containing both the CRD and the TMD, connected by the juxta-membrane linker domain (SMOΔC, Fig. 1a and Extended Data Fig. 1). To study the SMO TMD in a defined functional state and to reduce conformational flexibility, we included a single amino acid mutation, Val329Phe[16], in TMD helix 3 that locked SMO in an inactive state and substantially improved expression levels (Extended Data Fig. 2 and Supplementary Discussion). Using an established strategy in GPCR crystallography, the third intracellular loop (ICL3) between transmembrane helices 5 and 6 was replaced by thermostabilized apocytochrome *b*562RIL (BRIL)[21]. The SMOΔC structure was determined to 3.2 Å resolution (Extended Data Table 1 and Extended Data Fig. 3). The asymmetric unit, comprising two molecules arranged 'head-to-tail', stacks into alternating hydrophobic and hydrophilic layers along one axis, as is typical for lipidic cubic phase (LCP)-derived crystals (Extended Data Fig. 3a). This SMO arrangement within the crystal suggests that SMOΔC is monomeric, in agreement with size-exclusion chromatography (SEC) coupled to multi-angle light-scattering analysis (MALS) (Extended Data Fig. 3f).

SMO adopts an extended conformation in the structure. The extracellular CRD is perched on top of the linker domain, which forms a wedge between the TMD and CRD. At the apex of this wedge, the CRD contacts the TMD through the elongated TMD extracellular loop 3 (ECL3; Fig. 1a). The overall architecture is stabilized by nine disulfide bridges, four of which (numbered 2–5 in Fig. 1a) reveal the canonical disulfide pattern of the CRD fold[22] and one that is specific

[1]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [2]Departments of Biochemistry and Medicine, Stanford University School of Medicine, Stanford, California, 94305, USA. [3]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK. [4]Diamond Light Source Ltd, Harwell Science & Innovation Campus, Didcot OX11 0DE, UK. [5]Department of Developmental Biology, Washington University School of Medicine, St. Louis, Missouri, 63110, USA.
*These authors contributed equally to this work.

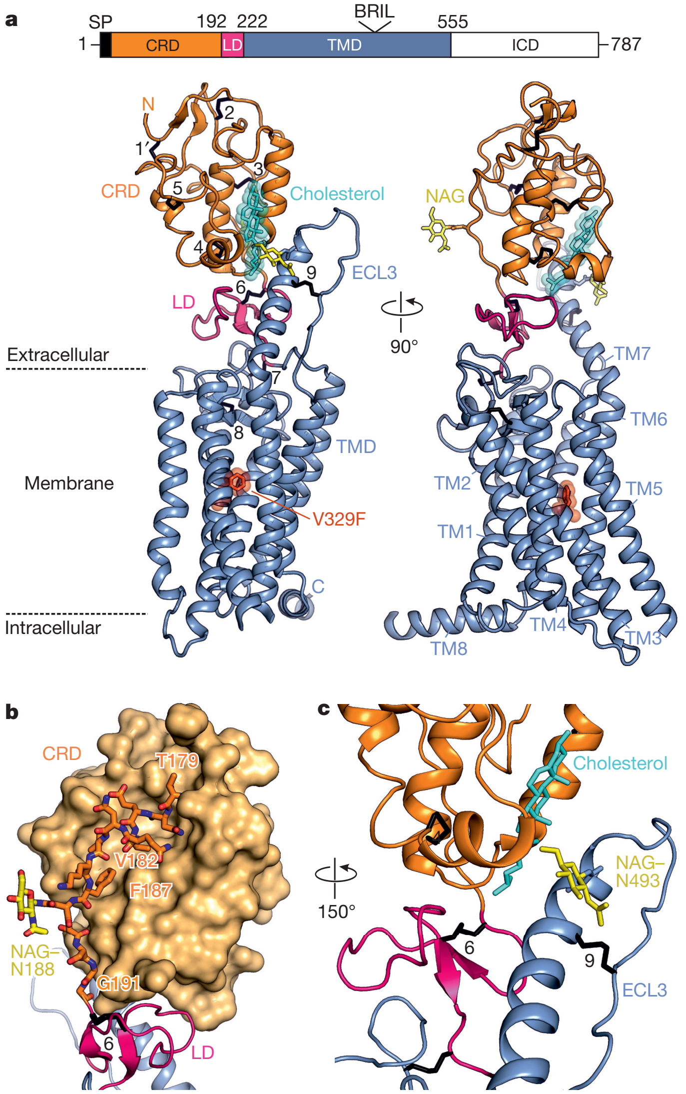


Figure 1 | Structure of human SMO. a, Two views of the overall structure showing the extracellular and transmembrane domains of human SMO in cartoon representation. Orange, CRD; pink, linker domain (LD); blue, TMD; red, inactivating point mutation Val329Phe; cyan, cholesterol; black, nine numbered disulfide bridges; yellow sticks, two *N*-linked glycans (NAG). A schematic of SMO is shown above (SP, signal peptide; BRIL, position of the BRIL fusion protein inserted between TMD helices 5 and 6). **b**, The 'connector' region between the CRD and linker domain highlighted as sticks in atomic colouring, with the CRD shown as a solvent-accessible surface and the linker domain and part of the TMD ECL3 loop as cartoons. **c**, Interface between the CRD, linker domain and TMD shown in cartoon representation. Yellow sticks, ECL3-NAG; cyan sticks, cholesterol.

for SMO[10] (marked as 1′ in Fig. 1a). This disulfide bridge positions the start of a 'connector' segment that links the CRD and linker domain (Fig. 1b). The connector is tucked along a hydrophobic groove that runs the length of the CRD and shields this groove with three inwardly turned hydrophobic residues (Val182, Ile185 and Phe187), multiple hydrophilic residues and an *N*-linked glycan facing the solvent (Fig. 1b and Extended Data Fig. 4a). The connector region, combined with mainly hydrophobic interactions between the CRD and both the linker domain and ECL3 (total buried surface area of 745 Å^2), orients the CRD in an upright conformation with its N terminus pointing away from the plasma membrane (Fig. 1a, b and Extended Data Fig. 4b). However, we did not observe major structural changes in the heptahelical TMD bundle when we compared it to previously solved structures of either antagonist- or agonist-bound complexes lacking the CRD[15–17]. The only exception was a rearrangement of the linker domain, which in our structure is pushed down towards the TMD, perhaps by CRD binding (Extended Data Fig. 5). A structure of the activated state of the SMO TMD will probably require co-crystallization with its (still unknown) downstream effector or the use of an active-state stabilizing antibody.

Cholesterol is a ligand for SMO

Unexpectedly, we discovered a cholesterol molecule in our SMO structure (Fig. 1 and Extended Data Methods). Cholesterol occupies a central position, interacting with all three SMO domains (the CRD, linker domain and TMD; Figs 1c and 2a, b), and adopts an extended conformation with its tetracyclic sterol ring bound in a shallow groove in the CRD, a site previously shown[10–12] to bind 20(S)-OHC in SMO and the palmitoleyl group of Wnt ligands in Frizzled receptors[23]. The cholesterol iso-octyl tail, located at the interface between the CRD, linker domain and TMD ECL3, is buried in the SMO protein core. This arrangement positions cholesterol some 12 Å away from the lipid bilayer of the plasma membrane (Fig. 2c) and indicates that a cholesterol molecule would have to completely desorb from the plasma membrane surface to access its binding site in SMO. Cellular cholesterol levels are permissive for Hh signalling[24,25]; however, this requirement is likely to be unrelated to the cholesterol binding site seen in our structure because SMO mutants lacking the CRD[12] or carrying mutations in the cholesterol-binding groove[25] remain sensitive to cholesterol depletion (Supplementary Discussion).

The sterol-binding site of SMO is predominantly lined with hydrophobic residues from the CRD, which stabilize the flat α-face of cholesterol (Fig. 2a). Mutations in several of these residues (Leu108, Trp109, Pro164 and Phe166) have been noted to prevent SMO binding to 20(S)-OHC and to impair signalling driven by either 20(S)-OHC or SHH[10–12]. The β-face is shielded by the side-chain of Arg161 and an *N*-linked glycan (Fig. 2b). The cholesterol 3β-hydroxyl group is

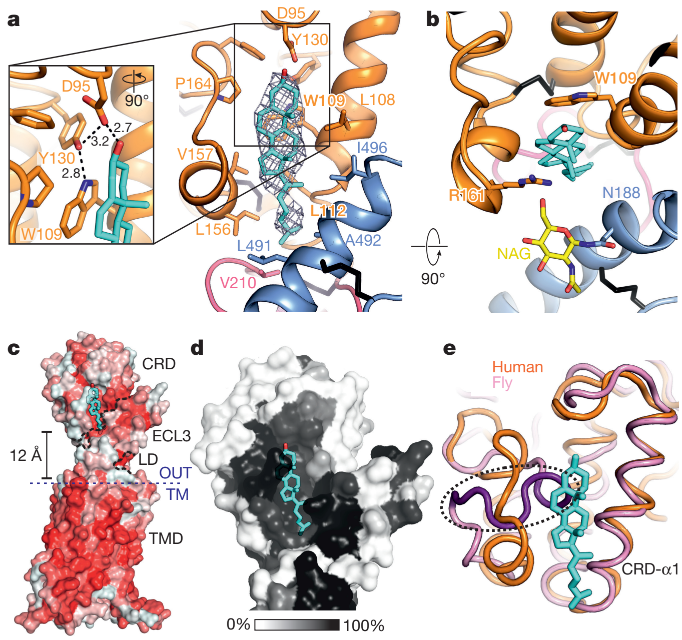


Figure 2 | The cholesterol binding site. a, b, Close-up of cholesterol with interacting residues as sticks. Initial $2F_o–F_c$ map at 1.0σ before inclusion of cholesterol shown as chicken-wire. Colour-coding follows Fig. 1. Inset shows the potential hydrogen-bonding network coordinating the cholesterol 3β-hydroxyl group. Interatomic distances (Å) are shown in black. **c**, Solvent-accessible surface colour-coded by hydrophobicity (red, hydrophobic; white, hydrophilic). **d**, Sequence conservation (based on 55 vertebrate SMO sequences) mapped onto SMOΔC (black, conserved; white, not conserved). **e**, Superposition of human (orange, this study) and fly (purple, PDB 2MAH[20]) SMO CRD structures. The fly CRD region occupying the cholesterol-binding site is highlighted by the dashed line.

incorporated in a hydrogen-bonding network with the side chains of Asp95, Trp109 and Tyr130; this optimally positions the α-face of the sterol ring system to make a stacking interaction with the indole ring of Trp109 (Fig. 2a, b). The Leu491, Ala492 and Ile496 residues of TMD ECL3 and the Val210 residue of the linker domain orient the iso-octyl tail of cholesterol in an elongated conformation. Residues lining the cholesterol-binding site are highly conserved in vertebrates but are less conserved in *Drosophila* (Fig. 2d and Extended Data Fig. 1). The superposition of our human CRD structure with a previously solved solution structure of the *Drosophila* CRD[20] (r.m.s.d. 1.71 Å for 86 equivalent Cα positions) revealed a major rearrangement of one edge of the cholesterol-binding site comprising fly SMO residues 183–190 (Fig. 2e and Extended Data Fig. 1). This segment of fly SMO forms a short helix that protrudes into the CRD groove and consequently may preclude cholesterol binding. In fact, fly SMO does not bind oxysterols[10,11].

Consistent with the structure, purified SMOΔC bound to cholesterol in a ligand-affinity assay (Fig. 3a–c), analogous to the assay we previously developed to measure the binding of SMO to 20(S)-OHC[7]. Beads covalently coupled to cholesterol captured purified SMOΔC (Fig. 3a). Binding could be blocked by free 20(S)-OHC added as a competitor at concentrations that activate Hh signalling in cells[7] (Fig. 3b), confirming that cholesterol and 20(S)-OHC engaged the same binding groove on the CRD surface. In a stringent specificity control, 20(R)-OHC, an epimer with inverted stereochemistry at a single position that cannot bind to the CRD[7] and cannot activate Hh signalling[10], failed to block this interaction (Fig. 3c).

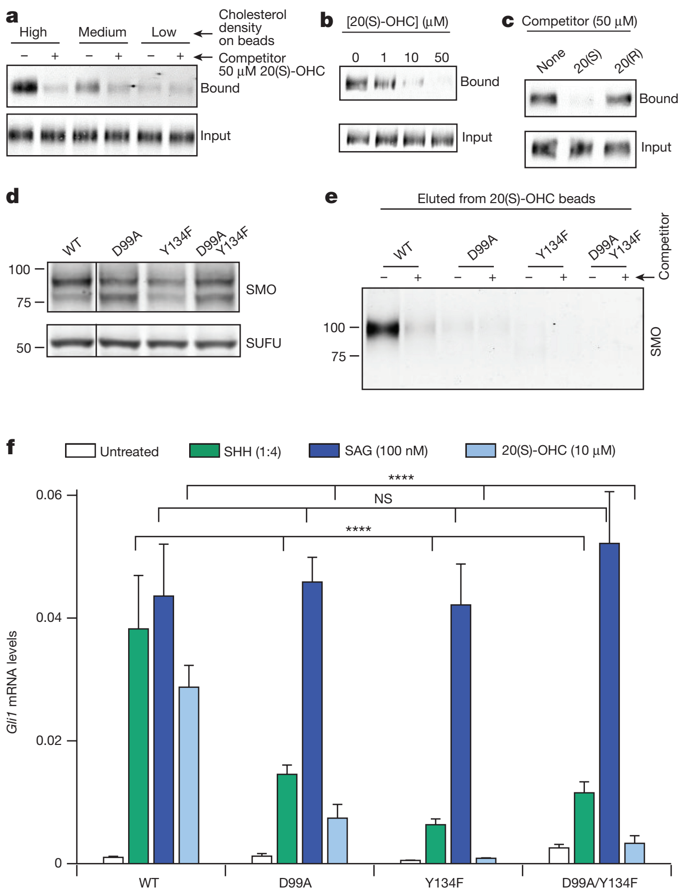


Figure 3 | The cholesterol-binding site regulates SMO signalling activity. **a**, Purified SMOΔC captured on beads coupled to increasing concentrations of cholesterol in the presence or absence of free 20(S)-OHC. **b**, **c**, SMOΔC captured on cholesterol beads in the presence of increasing concentrations of free 20(S)-OHC (**b**) or in the presence of a synthetic epimer 20(R)-OHC (**c**). **d**, Protein levels of SMO variants stably expressed in $Smo^{-/-}$ mouse fibroblasts. WT, wild-type. The protein Suppressor of Fused (SUFU) served as a loading control. **e**, Binding of SMO variants to 20(S)-OHC-coupled beads. Asp99Ala and Tyr134Phe are predicted to disrupt hydrogen bonding with cholesterol (Fig. 2a). **f**, Levels of *Gli1* mRNA (mean arbitrary units ± s.d., $n = 4$) were used as a metric for Hh signalling activity in cell lines shown in **d** after stimulation with SHH, SAG or 20(S)-OHC. Statistical significance based on one-way ANOVA is denoted for the difference in *Gli1* mRNA levels between cells expressing wild-type SMO and cells expressing each mutant SMO protein. n.s., $P > 0.05$; $^{****}P \leq 0.0001$. Each experiment was repeated 3 or more times.

Cholesterol promotes SMO signalling

To assess the functional relevance of the cholesterol-binding site identified in our structure, we focused our mutagenesis efforts on the hydrogen-bonding network between the cholesterol 3β-hydroxyl group and Asp95 and Tyr130 (Fig. 2a). The corresponding residues in mouse SMO, Asp99 and Tyr134, were mutated either individually or in combination to residues (Asp99Ala and Tyr134Phe) that lack hydrogen bond acceptor or donor groups. As cholesterol and oxysterols occupy the same groove in the CRD (Fig. 3b), they probably adopt a similar conformation and participate in at least some of the same interactions. Therefore, we used 20(S)-OHC binding and signalling to evaluate the effect of these mutations. First, we tested the ability of 20(S)-OHC beads to capture SMO from detergent extracts (Fig. 3d, e). As previously demonstrated[7], wild-type SMO can be captured on 20(S)-OHC beads. However, SMO variants carrying the Asp99Ala and Tyr134Phe point mutations failed to bind to 20(S)-OHC beads, highlighting the importance of this hydrogen-bonding network for sterol binding (Fig. 3e). Next, we stably expressed untagged versions of these SMO mutants in $Smo^{-/-}$ mouse fibroblasts (Fig. 3d) and assessed their abilities to restore signalling initiated by SHH, the TMD agonist SAG or the CRD agonist 20(S)-OHC (Fig. 3f). SMO-Asp99Ala and SMO-Tyr134Phe did not increase the basal activity of SMO in the absence of Hh agonists and did not significantly change the signalling response to SAG (Fig. 3f). This demonstrates that these mutants retained an intact TMD ligand-binding site and remained competent to transmit signals to cytoplasmic components. However, Asp99Ala and Tyr134Phe significantly impaired the ability of SMO to respond to both 20(S)-OHC and SHH (Fig. 3f). We conclude that these residues, and by implication the cholesterol seen in our structure, are important for signal-induced SMO activation.

Inactive-state stabilization by the SMO CRD

To assess the structural influence of the CRD in a membrane environment, we carried out molecular dynamics simulations of SMO in the presence and absence of cholesterol. Ten independent 100-ns atomistic simulations were performed with SMO embedded in a phosphatidylcholine bilayer (Fig. 4a, b and Extended Data Fig. 6). These simulations revealed that the SMO CRD has substantial conformational flexibility when not bound to a ligand. In the presence of cholesterol, however, there was a pronounced decrease in this flexibility (Fig. 4a), consistent with the idea that cholesterol stabilizes the CRD structure. By contrast, cholesterol did not substantially change the conformational stability of the SMO TMD (Fig. 4b). The predominant cholesterol-induced stabilization was seen in the vicinity of the sterol-binding pocket, with some propagation to more distal CRD regions (Extended Data Fig. 6). Consistent with these simulations, the thermostability of purified SMOΔC was reduced when cholesterol was depleted with methyl-β-cyclodextrin (Extended Data Fig. 6f, g and Supplementary Discussion).

The molecular dynamics results prompted us to test the signalling consequences of destabilizing interactions within the extracellular region of SMO. To probe the effect of interactions between the CRD and the linker domain, we generated two mutants (Pro120Ser and Ile160Asn/Glu162Thr) in the context of mouse SMO. These introduce bulky, *N*-linked glycosylation sites at positions 114 and 156, respectively, of the human CRD that contact the linker domain

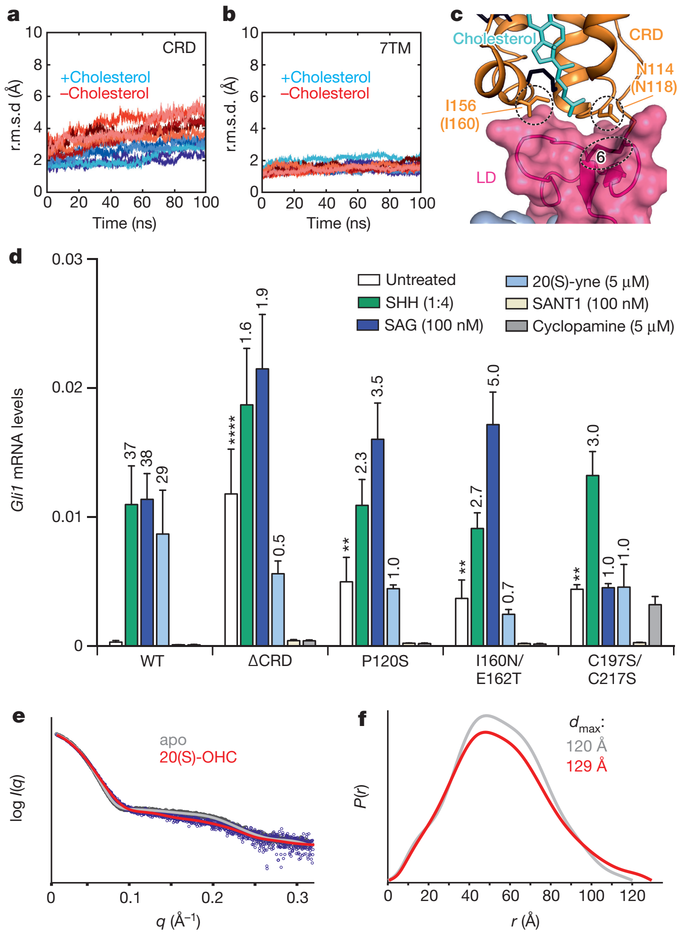


Figure 4 | SMO activity is regulated by the stability of its extracellular domain. a, **b**, Molecular dynamics (MD) simulations of SMO. Cα r.m.s.d. versus time for CRD (**a**) and 7TM region (**b**) with (blues) and without (reds) cholesterol. **c**, **d**, Mutations in the extracellular region increase constitutive signalling activity of SMO. **c**, CRD–linker domain interface with mutated residues highlighted. Corresponding mouse residues are in brackets. **d**, *Gli1* mRNA levels (mean arbitrary units ± s.d., $n = 4$, ≥ 3 independent repeats) after treatment with agonists (SHH, SAG, 20(S)-yne) or antagonists (cyclopamine, SANT-1). *Gli1* mRNA levels in untreated cells reflect constitutive activity of SMO. Numbers above bars indicate *Gli1* fold-increase compared to untreated cells. Asterisks denote statistical significance based on one-way ANOVA for comparison with wild-type SMO. $^{**}P \leq 0.01$, $^{****}P \leq 0.0001$. **e**, Solution-state SAXS profiles of untreated (grey) and 20(S)-OHC-treated (red) SMOΔC. Lines are fits derived from the indirect Fourier transform of the shown data points. **f**, Paired-distance ($P(r)$) distribution functions, normalized to their respective $I(0)$ values, with maximum particle sizes (d_{max}) of 120 Å for untreated (grey) and 129 Å for 20(S)-OHC-treated (red) SMOΔC.

(Fig. 4c). To disrupt the conformation of the linker domain itself, we mutated two cysteines (Cys197 and Cys217 in mouse SMO) that form a disulfide bond within the linker domain (numbered 6 in Figs 1a and 4c) to serines. We compared these mutants to SMO lacking the entire CRD (SMO-ΔCRD) (Fig. 4d and Extended Data Fig. 7a). Untagged SMO-ΔCRD stably expressed in *Smo*$^{-/-}$ fibroblasts demonstrated a similar level of constitutive activity[11,12] to that of wild-type SMO stimulated with saturating SHH (Fig. 4d and Extended Data Fig. 7b).

The Pro120Ser, Ile160Asn/Glu162Thr, and Cys197Ser/Cys217Ser mutations mimicked the effect of complete CRD deletion—constitutive signalling activity was increased (Fig. 4d) despite the presence of high levels of PTCH1 (Extended Data Fig. 7a) and there was complete loss of responsiveness to 20(S)-yne (Fig. 4d). All three mutants remained sensitive to inhibition by SANT-1, which binds deep in the TMD and does not contact the linker domain[16]. Cyclopamine and SAG, both of which make contacts with the linker domain[16,17], could regulate SMO-Pro120Ser and SMO-Ile160Asn/Glu162Thr, but not SMO-Cys197Ser/Cys217Ser. Thus, destabilization of either CRD–linker domain contacts or the linker domain itself increases the constitutive activity of SMO, implicating these regions in stabilizing an inactive state, analogous to the D(E)R$^{3.50}$Y motif in helix III of some Class A GPCRs[26].

To understand the basis for the loss of oxysterol responsiveness, we measured binding of each of the mutants to 20(S)-OHC beads (Extended Data Fig. 7c). Although the CRD mutations Pro120Ser and Ile160Asn/Glu162Thr impaired binding, the Cys197Ser/Cys217Ser mutation in the linker domain had no effect on binding to 20(S)-OHC. The observation that 20(S)-OHC cannot activate SMO-Cys197Ser/Cys217Ser (Fig. 4d) even though it can bind normally (Extended Data Fig. 7c) suggests that the linker domain may transmit the conformational changes that lead to SMO TMD activation in response to CRD ligands. To investigate conformational changes in SMO induced by CRD ligands in the solution state, we performed small angle X-ray scattering (SAXS) experiments on SMOΔC in the absence of any exogenous ligands (*apo*-SMOΔC) or SMOΔC loaded with the agonist 20(S)-OHC (Extended Data Fig. 8a). The initial SAXS curves (Fig. 4e) and further analyses[27,28] (Fig. 4f and Extended Data Fig. 8b) show that 20(S)-OHC binding induces a conformational change consistent with elongation and reduced globularity of SMOΔC. As 20(S)-OHC and cholesterol bind to the same groove on the CRD (Fig. 3b), we conclude that replacement of cholesterol by 20(S)-OHC (a molecule that carries only a single additional hydroxyl) produces a conformational change that leads to SMO activation.

The structure of SMO bound to vismodegib

Our structural and functional studies highlighted the critical regulatory role played by CRD–linker domain–TMD contacts in controlling the conformation and activity of SMO. It is unclear whether these interactions are altered by TMD-targeted small molecules used to treat patients with Hh-driven cancers, because prior structures of SMO bound to small molecules have not included the CRD, precluding an assessment of how the extracellular and transmembrane domains of SMO communicate. We determined the crystal structure of SMO in complex with vismodegib, a potent TMD antagonist, to 3.3 Å resolution (*vismo*–SMOΔC; Fig. 5a and Extended Data Fig. 9a, b). Despite the fact that vismodegib is the most commonly used Hh pathway inhibitor in patients and resistance has already become a clinically relevant problem[14,29,30], there are no structures available for vismodegib in complex with SMO.

Vismodegib is stabilized by a network of hydrogen bonds and hydrophobic interactions formed by the SMO TMD core (Fig. 5b). As expected, vismodegib occupies the TMD binding site previously identified in SMO complexes with other antagonists and agonists[15–17] (Extended Data Fig. 5). The vismodegib pyrimidine and chloro-benzyl rings are deeply buried, forming mainly hydrophobic interactions with the TMD core. The amide linker, which is stabilized by a potential hydrogen bond to the side chain of Asp384, occupies a central position within the seven transmembrane domain (7TM) bundle and connects to the chlorophenyl–methylsulfone moiety, which is oriented towards the extracellular domains and the entrance of the TMD pocket[16]. This arrangement is stabilized by potential hydrogen bonds to the side chains of Gln477 and Arg400 and a hydrophobic stacking interaction of the vismodegib methylsulfone moiety and the aromatic ring of Phe484. Asp473, a residue that is mutated in vismodegib-resistant cancers[14,29,30], stabilizes the potential hydrogen-bonding network around Arg400 (asterisk in Fig. 5b). The SMO-Asp473His mutation seen in drug-resistant cancers is ideally positioned to disrupt the vismodegib-binding site, in agreement with functional experiments showing that SMO-Asp473His cannot bind this drug[14]. Our structure also explains the vismodegib resistance mechanism for many other mutations in the TMD binding site observed in patients with advanced basal cell cancer[29,30] (Extended Data Fig. 9c–f).

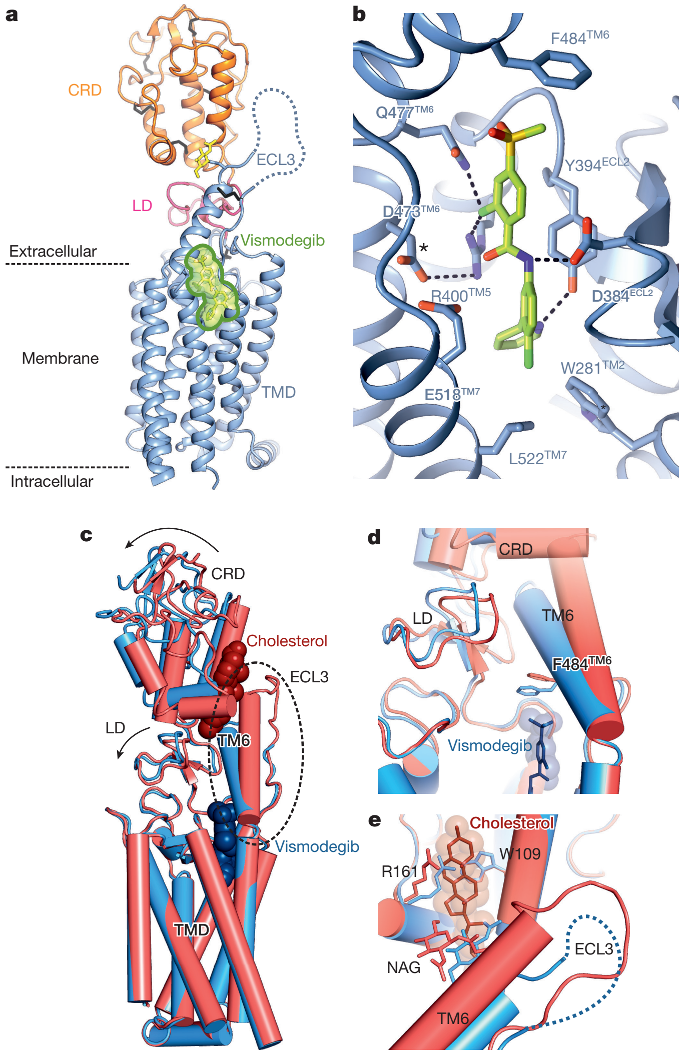


Figure 5 | Structure of SMO in complex with the antagonist vismodegib. **a**, Overall structure showing full extracellular and transmembrane domains of human SMO (cartoon) in complex with vismodegib (green). Colour-coding follows Fig. 1. **b**, Close-up of the vismodegib-binding site. The asterisk denotes a residue mutated in vismodegib-resistant tumours. **c**–**e**, Comparison of *apo*-SMO (red) and *vismo*–SMO (blue). **c**, Superposition of the two SMO structures based on the TMD. Arrows indicate domain rotations. Cholesterol (red) and vismodegib (blue) shown as spheres. The dashed oval highlights the conformational change of the TM6 and ECL3. **d**, Close-up of the TMD–linker domain–CRD interface. Phe484 contacts the vismodegib methylsulfone group. **e**, Close-up of the sterol-binding site. Cholesterol, Arg161 and NAG shown as sticks. The partly disordered ECL3 loop of *vismo*–SMO is depicted as dotted line.

The vismodegib-bound structure of SMO in its inactive state shows a marked conformational change in the extracellular domains and the ECL3 loop when compared to the *apo*-SMO structure (Fig. 5c). The upper part of transmembrane helix 6, which contacts the CRD and linker domain, is moved ~15° towards the linker domain and CRD, probably owing to an interaction between Phe484 and the methylsulfone moiety of vismodegib (Fig. 5b, d). This movement of helix 6 results in a rotational movement of the CRD (and to a lesser extent the linker domain), allowing ECL3 to intrude into the sterol-binding groove of the CRD. As a consequence, the side chain of CRD residue Arg161 forms a stacking interaction with Trp109 and occupies the space where cholesterol was located in the *apo*-SMO structure (Fig. 5e). This reorganization is predicted to occlude the cholesterol-binding site. Indeed, cholesterol was absent from the *vismo*–SMO structure (Fig. 5a), and the addition of vismodegib reduced the interaction of purified SMOΔC with cholesterol beads (Extended Data Fig. 9g, h). In summary, binding of vismodegib to the TMD site induces a conformational change that ultimately results in rearrangement of the sterol-binding site, providing a structural communication mechanism that explains the previously observed allosteric interaction[7] between CRD and TMD ligands. More generally, this finding suggests that a similar conformational change, involving a shift of the ECL3 and a pivoting of the CRD on the extracellular end of the TMD bundle, may allow an extracellular signal to be transduced from the CRD to the TMD and ultimately across the membrane. From a therapeutic perspective, our results highlight an unanticipated role for the CRD and ECL3, including displacement of an extracellular cholesterol ligand, in stabilizing a drug-bound inactive state of SMO, providing a structural template for the development of the next generation of SMO inhibitors against Hh-driven cancers.

Conclusion

The structure of SMO provides insights into the mechanism by which a large extracellular region and two allosterically linked ligand-binding sites may regulate the activity of a GPCR. We propose that cholesterol functions as an endogenous SMO ligand that occupies the CRD groove and stabilizes a resting or *apo* conformation poised to respond to Hh signals. SMO signalling activity is compromised by mutations that prevent cholesterol binding or by antagonists such as vismodegib that act allosterically to occlude the cholesterol-binding site. SAXS data suggest that CRD ligands such as 20(S)-OHC, which displace cholesterol, produce an additional conformational change that leads to SMO activation. Identification of the mechanism by which PTCH1 inhibits SMO will be necessary to understand how cholesterol-bound SMO is activated in response to endogenous Hh morphogens. We predict that SMO activation will involve alterations in the interactions between the CRD, ECL3, linker domain, and TMD that allow the TMD to adopt an active signaling state.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 November 2015; accepted 15 June 2016.
Published online 20 July 2016.

1. Sharpe, H. J., Wang, W., Hannoush, R. N. & de Sauvage, F. J. Regulation of the oncoprotein Smoothened by small molecules. *Nat. Chem. Biol.* **11,** 246–255 (2015).
2. Taipale, J. *et al.* Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* **406,** 1005–1009 (2000).
3. Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* **16,** 2743–2748 (2002).
4. Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. Small molecule modulation of Smoothened activity. *Proc. Natl Acad. Sci. USA* **99,** 14071–14076 (2002).
5. Frank-Kamenetsky, M. *et al.* Small-molecule modulators of Hedgehog signaling: identification and characterization of Smoothened agonists and antagonists. *J. Biol.* **1,** 10 (2002).
6. Robarge, K. D. *et al.* GDC-0449-a potent inhibitor of the hedgehog pathway. *Bioorg. Med. Chem. Lett.* **19,** 5576–5581 (2009).
7. Nachtergaele, S. *et al.* Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nat. Chem. Biol.* **8,** 211–220 (2012).
8. Corcoran, R. B. & Scott, M. P. Oxysterols stimulate Sonic hedgehog signal transduction and proliferation of medulloblastoma cells. *Proc. Natl Acad. Sci. USA* **103,** 8408–8413 (2006).
9. Dwyer, J. R. *et al.* Oxysterols are novel activators of the hedgehog signaling pathway in pluripotent mesenchymal cells. *J. Biol. Chem.* **282,** 8959–8968 (2007).
10. Nachtergaele, S. *et al.* Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. *eLife* **2,** e01340 (2013).
11. Nedelcu, D., Liu, J., Xu, Y., Jao, C. & Salic, A. Oxysterol binding to the extracellular domain of Smoothened in Hedgehog signaling. *Nat. Chem. Biol.* **9,** 557–564 (2013).
12. Myers, B. R. *et al.* Hedgehog pathway modulation by multiple lipid binding sites on the smoothened effector of signal response. *Dev. Cell* **26,** 346–357 (2013).
13. Briscoe, J. & Thérond, P. P. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* **14,** 416–429 (2013).
14. Yauch, R. L. *et al.* Smoothened mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science* **326,** 572–574 (2009).

15. Wang, C. *et al.* Structure of the human smoothened receptor bound to an antitumour agent. *Nature* **497,** 338–343 (2013).
16. Wang, C. *et al.* Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nat. Commun.* **5,** 4355 (2014).
17. Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5,** 3309 (2014).
18. Deupi, X. & Standfuss, J. Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* **21,** 541–551 (2011).
19. Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53,** 531–556 (2013).
20. Rana, R. *et al.* Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling. *Nat. Commun.* **4,** 2965 (2013).
21. Chun, E. *et al.* Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20,** 967–976 (2012).
22. Bazan, J. F. & de Sauvage, F. J. Structural ties between cholesterol transport and morphogen signaling. *Cell* **138,** 1055–1056 (2009).
23. Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science* **337,** 59–64 (2012).
24. Cooper, M. K. *et al.* A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis. *Nat. Genet.* **33,** 508–513 (2003).
25. Blassberg, R., Macrae, J. I., Briscoe, J. & Jacob, J. Reduced cholesterol levels impair Smoothened activation in Smith-Lemli-Opitz syndrome. *Hum. Mol. Genet.* **25,** 693–705 (2016).
26. Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494,** 185–194 (2013).
27. Durand, D. *et al.* NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169,** 45–53 (2010).
28. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95,** 559–571 (2011).
29. Sharpe, H. J. *et al.* Genomic analysis of smoothened inhibitor resistance in basal cell carcinoma. *Cancer Cell* **27,** 327–341 (2015).
30. Atwood, S. X. *et al.* Smoothened variants explain the majority of drug resistance in basal cell carcinoma. *Cancer Cell* **27,** 342–353 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff of beamlines I24 and B21 (MX10627) at the DLS, UK for assistance. We thank S. Masiulis, D. Staunton, K. Jungnickel, A. R. Aricescu and G. Schertler for discussions. The work was supported by Cancer Research UK (C20724/A14414), the US National Institutes of Health (GM106078, HL067773), the Wellcome Trust (102890/Z/13/Z, 092970/Z/10/Z and 090532/Z/09/Z), and the Taylor Family Institute for Psychiatric Research. Further support by NDM Oxford (E.F.X.B.), Medical Research Council UK (G.H.), Ford Foundation (G.L.) and National Science Foundation (S.Na.) is acknowledged.

Author Contributions E.F.X.B. produced the protein with P.S.M, and carried out crystallization with S.Ne. C.S. and E.F.X.B. determined the crystal structures. R.S., S.Na., and G.L. performed Hh signalling and biochemical assays. L.M.-M. and D.F.C. synthesized sterol analogues. G.H. and M.S.P.S performed MD analysis. M.D.T and R.P.R. carried out SAXS analysis. C.S. and R.R. supervised the project. E.F.X.B., R.R. and C.S. wrote the paper, with input from all authors.

Author Information Atomic coordinates and structure factors for the *apo*-SMOΔC and *vismo*–SMOΔC crystal structures have been deposited in the Protein Data Bank (PDB) under accession numbers 5L7D and 5L7I. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.R. (rrohatgi@stanford.edu) or C.S. (christian@strubi.ox.ac.uk).

Reviewer Information *Nature* thanks J. Briscoe, R. Dror, F. de Sauvage and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Reagents. NIH 3T3 and 293T cells were certified stocks obtained directly from ATCC. *Smo*$^{-/-}$ fibroblasts (which were used to express all the SMO mutants) have been described previously[31] and were originally obtained from J. Chen and P. Beachy. The authenticity of *Smo*$^{-/-}$ cells was established by immunoblotting to ensure lack of endogenous SMO protein expression. Incoming cell lines were confirmed to be negative for mycoplasma contamination. SAG was obtained from Enzo Life Sciences; cyclopamine was obtained from Toronto Research Chemicals; SANT-1 was obtained from EMD Millipore; 20(S)-OHC was obtained from Steraloids. The synthesis of 20(S)-yne, 20(R)-OHC[7] and 20(S)-amine-coupled beads for SMO binding assays has been described in detail previously[10]. Rabbit polyclonal antibodies against SMO, PTCH1 and SUFU have been described previously[32,33], and the mouse monoclonal antibody against GLI1 was obtained from Cell Signaling Technologies (L42B10). SHH-containing conditioned medium was made from 293T cells transfected with the N-terminal signalling domain of SHH and used at saturating concentrations (dilution of 1:4)[2].

Constructs. For large-scale expression and crystallization, a SMO construct (SMOΔC) was designed by truncating the N and C termini of human SMO to leave the extracellular and transmembrane domains (UniProt Id. Q99835; residues 32–555), and replacing intracellular loop 3 (Q99835; residues 429–440) with BRIL (UniProt Id. P0ABE7; residues 23–128). The synthetic gene encoding SMOΔC was obtained from Geneart (Grand Island, NY) and cloned into the pHLSec vector[34] in frame to either a C-terminal Rho1D4 antibody epitope tag[35,36] or a monoVenus tag[37,38] followed by a Rho1D4 antibody epitope tag.

All mouse SMO mutants were made using Quikchange methods using a previously described construct[39] encoding mouse SMO (pCS2+:YFP-mSmo), after removal of the N-terminal YFP tag by XhoI digestion. The mouse SMO-ΔCRD construct lacks residues 68–184. For stable-line construction, the mouse SMO coding sequence was transferred from pCS2+ to pMSCVpuro using Gibson cloning.

Expression and purification of SMO. SMOΔC was expressed by transient transfection in HEK-293S-GnTI^{-} (ATCC CRL-3022) cells in a typical batch volume of 9.6 l. Cells were grown in suspension at 37 °C, 8.0% CO_2, 130 r.p.m. to densities of $2–3 \times 10^6$ cells ml^{-1} in protein expression medium (PEM, Invitrogen) supplemented with L-glutamine, non-essential amino-acids (NEAA, Gibco) and 1% fetal calf serum (FCS, Sigma-Aldrich). Cells from 0.8 l cultures were collected by centrifugation (1,100 r.p.m., 7 min) and re-suspended in 120 ml Freestyle293 medium (Invitrogen) containing 1.2 mg PEI Max (Polysciences), 0.4 mg plasmid DNA and 5 mM valproic acid (Sigma-Aldrich), followed by a 3–6-h incubation at 160 r.p.m. Culture media were subsequently topped up to 0.8 l with PEM and returned to 130 r.p.m. 48–72 h after transfection, cell pellets were collected by centrifugation, snap-frozen in liquid N_2 and stored at −80 °C, resulting in a total of ~150 g of cell mass per 9.6 l suspension medium[40].

Frozen cell pellets were thawed, re-suspended in 10 mM HEPES pH 7.5, 300 mM NaCl buffer supplemented with a 1:100 (v:v) dilution of mammalian protease inhibitor cocktail (P8340, Sigma-Aldrich) and solubilised with 1.3% (w/v) *n*-dodecyl-β-D-maltopyranoside (DDM, Anatrace) and 0.26% (w/v) cholesteryl hemisuccinate (CHS, Anatrace), then rotated for 1.5 h at 4 °C. Insoluble material was removed by centrifugation (10,000 r.p.m., 12 min, 4 °C) and the supernatant incubated for 2 h at 4 °C with purified Rho-1D4 antibody (University of British Columbia) coupled to CNBr-activated sepharose beads (GE Healthcare). Protein-bound beads were washed extensively with 50 mM HEPES pH 7.5, 300 mM NaCl, 10% (v/v) glycerol, 0.1% DDM, 0.02% CHS buffer and then with 50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 0.05% DDM, 0.01% CHS buffer and eluted overnight in 50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS, 500 μM TETSQVAPA peptide (Genscript). Eluate was concentrated to ~500 μl using a Vivaspin Turbo 4 PES 100 kDa MWCO concentrator and loaded onto a Superose 6 10/300 column (GE Healthcare) equilibrated with 10 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol, 0.03% DDM, 0.006% CHS. Peak fractions were pooled and concentrated to ~30 mg ml^{-1} using a Vivaspin 500 PES 100 kDa MWCO concentrator. Samples were deglycosylated with Endoglycosidase F1 and incubated at room temperature for 1 h. For the vismodegib complex, vismodegib (GDC-0449, Selleck Chem) dissolved at high concentration in DMSO was added to the protein sample to a final concentration of 10 mM.

For small-scale screening, cells in adherent format were transiently transfected using Lipofectamine2000 (Invitrogen) and the expressed protein, tagged with YFP[37,38], was prepared in the same manner as above with quantities adjusted appropriately. For analysis, samples were loaded onto a Superose 6 3.2/300 column (GE Healthcare) attached to a high-performance liquid chromatography system with automated micro-volume loader and in-line fluorescence detection (Shimadzu)[41].

For thermostability experiments, SMOΔC was expressed and purified as described above. After SEC purification, the pooled peak fractions were re-applied to purified Rho-1D4 antibody coupled to CNBr-activated sepharose beads. Equal amounts of beads were treated with different quantities of methyl-β-cyclodextrin (MBCD) and incubated with gentle rocking for 1 h at 15 °C before extensive washing with 300 mM NaCl, 50 mM Hepes pH 7.5, 0.03% DDM and elution with 300 mM NaCl, 50 mM Hepes, pH 7.5, 0.03% DDM, 500 μM TETSQVAPA peptide. The thermostability of the different samples was assessed by heating aliquots of their eluates to the indicated temperatures for 1 h before loading them onto a Superose 6 3.2/300 column on a Shimadzu system and following using absorbance at 280 nm. Samples were kept at the baseline temperature of 20 °C when not heated. The construct used in this assay was not fluorescently tagged in order to avoid the potentially confounding effects of the fluorescent tag on overall stability.

Stable cell lines. Stable cell lines expressing untagged SMO mutants were made by infecting *Smo*$^{-/-}$ fibroblasts with a retrovirus carrying these constructs cloned into pMSCVpuro[39]. Retroviral supernatants were produced after transient transfection of Bosc23 helper cells with the pMSCV constructs[42,43]. Virus-containing media from these transfections was directly used to infect *Smo*$^{-/-}$ fibroblasts, and stable integrants were selected with puromycin (2 μg ml^{-1}).

We had previously[10] constructed stable lines using an identical strategy with SMO constructs carrying an N-terminal fluorescent protein (FP) tag; however, we found that epitope tagging of SMO with a fluorescent protein, or transient overexpression of tagged or untagged SMO, could impact the assessment of its signalling activity. For example, we previously measured a lower level of constitutive signalling activity for SMO-ΔCRD compared to the present study, probably due to the presence of an N-terminal YFP tag[10]. Moreover, previous reports from three groups (including our own) reached somewhat divergent conclusions regarding the role of the CRD in basal and ligand-stimulated SMO activity[10–12]. These differences may have been related to the use of different epitope tags and expression systems.

Hence, all Hh signalling assays used in this study were performed with untagged SMO and SMO mutants stably expressed in *Smo*$^{-/-}$ fibroblasts, with assessment of SMO protein levels in stable cell lines by immunoblotting.

Hedgehog signalling assays. Stable cell lines expressing either wild-type SMO or SMO mutants were grown to confluence in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% Fetal Bovine Serum (FBS, Optima Grade, Atlanta Biologicals) and then exposed to DMEM containing 0.5% FBS for 24 h to induce primary cilia assembly. These ciliated cells were then treated with saturating concentrations of various Hh agonists and antagonists in DMEM containing 0.5% FBS for 12 h.

For detection of proteins by immunoblotting, cells were washed in ice-cold PBS and lysed (30 min, 4 °C) by agitation in modified RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM sodium chloride, 2% NP-40, 0.25% deoxycholate, 0.1% sodium-dodecyl sulfate (SDS), 1 mM dithiothreitol, 1 mM sodium fluoride and the SigmaFast Protease inhibitor cocktail). After clarification (20,000*g*, 30 min, 4 °C), the protein concentration of each lysate was measured using the bicinchoninic acid assay (BCA, Pierce/Thermo Scientific). Lysate aliquots containing equal amounts of total protein were fractionated on SDS–PAGE gels (either a 8% tris-glycine gel or a 4–12% bis-tris gel), and transferred to a nitrocellulose membrane. Quantitative immunoblotting with the various antibodies was performed using the Li-Cor Odyssey infrared imaging system. In all immunoblots, vertical dashed black lines represent non-contiguous lanes from the same immunoblot juxtaposed for clarity. Each immunoblot was repeated 2–3 times for all experiments shown.

Gli1 mRNA is a commonly used metric for Hh signalling activity, because *Gli1* is a direct Hh target gene. *Gli1* and *Gapdh* mRNA levels were measured by quantitative, reverse-transcription PCR (qRT–PCR) using the Power SYBR Green Cells-To-CT kit from Thermo Fisher Scientific and custom primers for *Gli1* (forward primer: 5′-CCAAGCCAACTTTATGTCAGGG-3′ and reverse primer: 5′-AGCCCGCTTCTTTGTTAATTTGA-3′) and *Gapdh* (forward primer: 5′-AGTGGCAAAGTGGAGATT-3′ and reverse primer: 5′-GTGGAGTCATACTGGAACA-3′). Transcript levels relative to Gapdh were calculated using the ΔC_t method and reported in arbitrary units. Each qRT–PCR experiment, which was repeated 2–4 times, included two biological replicates, each with two technical replicates.

Statistical analysis of *Gli1* mRNA levels across samples was performed using an ordinary one-way ANOVA test with a Holm–Sidak post-test to correct for multiple comparisons using the GraphPad Prism suite. Statistical significance in the figures is denoted as follows: NS, $P > 0.05$; $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $****P \leq 0.0001$.

Ligand affinity chromatography. Ligand affinity chromatography to assess the interaction between SMO protein in detergent extracts and beads covalently coupled to 20(S)-amine has been described previously[10]. Membranes from cells transiently or stably expressing constructs encoding mouse SMO variants were lysed in a DDM extraction buffer (50 mM Tris pH 7.4, 150 mM NaCl,

10% v/v glycerol, 0.5% w/v DDM and the SigmaFast EDTA-free protease inhibitor cocktail) for 2 h at 4 °C, followed by removal of insoluble material by ultracentrifugation (100,000*g*, 30 min). This DDM extract was incubated with 20(S)-OHC beads overnight at 4 °C to allow binding to equilibrium. After extensive washing, proteins captured on the beads were eluted with reducing LDS sample buffer (Life Technologies) and 100 mM dithiothreitol. The presence of SMO in these eluates was determined by quantitative immunoblotting with an anti-SMO antibody[32] and infrared imaging (Li-Cor Odyssey). Each experiment was repeated twice with similar results.

Preparation of cholesterol–PEG3–sepharose using the azide-alkyne Huisgen cycloaddition reaction from the Click Chemistry toolbox. 100 μl packed PEG3-azide sepharose resin (22 μmole per ml, Click Chemistry Tools) was washed 3 × 1 ml with 20% ethanol (v/v aq.) and re-suspended in 500 μl 20% ethanol (v/v). The bead slurry was supplemented with 1 mM $CuSO_4$, 5 mM Tris (3-hydroxypropyltriazolylmethyl)amine (THPTA), and 15 mM sodium ascorbate. The cycloaddition coupling reaction was initiated by adding 0.044 μmoles (high) of LKM-26, a previously described[44,45] alkynyl cholesterol derivative synthesized in-house, to achieve an approximate ligand density of 1:50 (moles coupled azide functional groups: moles uncoupled azide functional groups). High, medium and low ligand densities (Fig. 2f) represent coupling ratios of 1:50, 1:200 and 1:1,000. The reaction was protected from light and allowed to proceed at room temperature for 20 h with end-over-end rotation. The supernatant was removed from the resin, and the reaction quenched with 1 mM EDTA in 20% ethanol (v/v aq.). The supernatant was extracted with diethyl ether and loaded on a normal-phase TLC plate next to an alkynyl cholesterol standard to assess the efficiency of coupling. The thin layer chromatography (TLC) plate was developed using 2% methanol in chloroform (v/v) and stained using 10% $CuSO_4$/10% H_3PO_4 followed by charring at 200 °C for visualization.

SMO pull-down assays using cholesterol–PEG3–sepharose. Binding reactions were carried out in buffer containing 20 mM HEPES pH 7.4, 150 mM NaCl, and 0.03% DDM in a final volume of 100 μl. Each reaction contained 1 μg purified SMOΔC, the same protein used for crystallization studies. Competitors were added to the binding reaction and incubated for 1 h at room temperature before the addition of 20 μl packed cholesterol–PEG3–sepharose. Reactions were incubated for 16 h at 4 °C for affinity capture. The binding reactions were washed 3 × 1 ml with binding buffer and eluted using 2× Laemmli buffer for 30 min at room temperature. SMOΔC levels in the input, flow-through and captured on the beads were determined by quantitative immunoblotting (Li-Cor Odyssey) using the 1D4 primary antibody (mouse 1:2,000). Each experiment was repeated twice with similar results.

Crystallization and data collection. Protein samples were reconstituted into lipidic cubic phase (LCP) by mixing with molten lipid in a mechanical syringe mixer[46]. Molten lipid, consisting of 10% cholesterol (Sigma-Aldrich) and 90% 9.9 monoacylglycerol (monoolein, Sigma-Aldrich), was mixed with detergent-solubilized protein (either *apo* or with 10 mM vismodegib in DMSO) at ~30 mg ml^{-1} in a ratio of 60:40. A Gryphon robot (Art Robbins Instruments) was used to dispense 50 nl boluses of protein laden mesophase followed by 0.8 μl of precipitant solution onto each of 96 positions on a siliconized glass plate, which were then covered with a coverslip in a 'sandwich-plate' format. Crystals were grown at 20 °C and monitored by eye, using a microscope fitted with cross-polarizers, and subsequently imaged using a UV-imaging system (Rigaku Minstrel). *apo*-SMOΔC crystallized in 0.1 M MES pH 6, 30% (v/v) PEG500 DME, 0.1 M sodium acetate, 0.5 mM zinc chloride, 0.1 M ammonium fluoride. *vismo*-SMOΔC crystallized in 0.09 M sodium acetate pH4, 0.09 M sodium malonate, 27% (v/v) PEG500 DME, 0.1 M sodium acetate, 0.5 mM zinc chloride, 0.1 M ammonium fluoride.

X-ray data collection was conducted at MX beamline I24 at the Diamond Light Source (Harwell, UK). Prior to data collection crystals were flash-frozen at 105 K. X-ray data were processed using Xia2 (refs 47, 48), scaled using XSCALE[48] and merged using Aimless[49,50]. The final data set used for structure solution and refinement was merged from data from nine crystals for *apo*-SMOΔC and two crystals for *vismo*-SMOΔC. Data collection statistics are shown in Extended Data Table 1.

Structure determination, refinement and analysis. The *apo*-SMOΔC structure was solved by molecular replacement in PHASER[51] using the structure of human SMO TMD (PDB 4QIM[16]), zebrafish SMO CRD (PDB 4C79[10]) and BRIL (PDB 4EIY[52]) as search models. Extra electron density accounting for the region between CRD and LD, BRIL and TMD was immediately discernible after density modification in PARROT[53] (Extended Data Fig. 3e). We also observed extra density within the CRD ligand binding pocket (Fig. 2a) and assigned this to cholesterol bound in a stereo-specific orientation based on shape, coordination and refinement statistics (cholesterol addition improved the *R*-factors by over 1%), which is also in agreement with the markedly lower *B*-factor of the refined cholesterol compared to the protein backbone. Cholesterol may be derived from the cells (mM concentrations within the cell membrane) or from the LCP crystallization mix (that contained 10% (w/v) cholesterol). The *apo*-SMOΔC polypeptide chain was traced using iterative rounds of BUCCANEER[54], manual building in COOT[55] and refinement in autoBUSTER[56] and PHENIX[57]. This resulted in a well-defined model for the *apo*-SMOΔC structure that included two molecules of SMO (residues 59–549) with a BRIL protein segment inserted between SMO TMD helices 5 and 6, two *N*-linked glycans and a cholesterol molecule. We observed a systematic disorder along the *c* axis (Extended Data Fig. 2a), resulting in alternating ordered and less ordered hydrophilic layers within the LCP-grown crystals. This was not caused by crystal non-isomorphy or pseudo-symmetry, because reducing the space group from *C*2 to *P*1 had no effect on the disorder. The *vismo*–SMOΔC structure was solved by molecular replacement using the *apo*-SMOΔC structure. Extra electron density accounting for vismodegib and two well-ordered monoolein molecules was immediately apparent. The structure was refined using autoBUSTER[56] and PHENIX[57] with non crystallographic and secondary structure restraints. Crystallographic and Ramachandran statistics are given in Extended Data Table 1. Stereochemical properties were assessed by MOLPROBITY[58]. Superpositions were calculated using the program COOT[55], electrostatic potentials were generated using APBS[59] and hydrophobicity was calculated according to the Eisenberg hydrophobicity scale[60], as implemented in PyMOL[61]. Buried surface areas of protein–protein interfaces were calculated using the PISA webserver[62] with a probe radius of 1.4 Å. Sequence alignment was performed using MULTALIN[63] and formatted with ESPRIPT[64]. Program Caver was used with default settings to visualize the SMO TMD ligand binding pocket[65].

Amphipol exchange and MALS. In order to avoid background light-scattering due to free detergent in solution, protein samples were exchanged into amphipol[66](A8-35, Anatrace) at a mass ratio of 3:1 amphipol:protein and rotated at room temperature for 30 min. BioBeads (BIORAD), equilibrated in detergent-free buffer (10 mM HEPES pH 7.5, 300 mM NaCl), were added to the protein–detergent–amphipol mixture at 10 mg per 100 μl and incubated overnight at 4 °C to remove all detergent molecules. For multi-angle light scattering (MALS) experiments, amphipol-solubilized protein at 1 mg ml^{-1} was loaded onto a Superose 6 10/300 column (GE Healthcare), equilibrated in detergent-free buffer, on a Shimadzu system with inline MALS detector (Wyatt). Data were analysed using ASTRA6.1.2 software (Wyatt). For SMOΔC, the values used for d*n*/d*c* and ε at 280 nm were 0.185 ml g^{-1} and 1.541 ml (mg.cm)$^{-1}$, respectively. For protein conjugate analysis, the d*n*/d*c* used for amphipol A8-35 was 0.15 ml g^{-1} (ref. 67).

Molecular dynamics system setup. Simulations were performed using the GROMACS v4.6.3 simulation package[68]. Side chain ionization states were modelled using pdb2gmx (Histidine) and PropK_a (all other residues)[69,70]. The N and C termini were treated with neutral charge. Intracellular loop 3 (occupied by the BRIL fusion in our crystal structure) was modelled using coordinates from the PDB entry 4N4W (ref. 16). The protein structure was then energy-minimized using the steepest descents algorithm implemented in GROMACS, before being converted to a coarse-grained representation using the MARTINI 2.2 force field[71]. The energy-minimized coarse-grained structure was centred in a simulation box with dimensions 100 × 100 × 180 Å^3. 270 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipids were randomly placed around the protein and the system solvated and neutralised to a concentration of 0.15 M NaCl. An initial 1 μs of coarse-grained simulation was applied to permit the self-assembly[72] of a POPC lipid bilayer around the GPCR. During the coarse-grained simulation, the structure of the protein was maintained by an elastic network, allowing local conformational flexibility of the protein. Thus, the protein was able to adopt its optimal orientation within the lipid bilayer[73]. The endpoint of the coarse-grained bilayer self-assembly simulation was converted back to atomic detail using a fragment-based protocol for the lipid conformations[74], while retaining the original crystal structure of the protein, now located in its optimal orientation and position within the lipid bilayer. Equilibration of the atomic system was achieved through 1 ns of NPT simulation with the protein coordinates restrained, before the system was subjected to 100 ns of unrestrained atomistic molecular dynamics. Simulations were performed both in the presence and absence of the cholesterol ligand. Five repeat simulations were run for each case.

Coarse-grained simulations. The standard MARTINI force field[75] and its extension to proteins[71,76] was used to describe all system components. During the coarse-grained self-assembly simulation an ELNEDYN network[77] was applied to the protein using force constant of 500 kJ mol^{-1} nm^{-2} and a cutoff of 1.5 nm. Temperature was maintained at 310 K using a Berendsen thermostat[78] with a coupling constant of $\tau_t = 1$ ps, and pressure was controlled at 1 bar using a Berendsen barostat[78] with a coupling constant of $\tau_p = 1$ ps and a compressibility of 5×10^{-6} bar^{-1}. Electrostatics and van der Waals interactions in the CG simulations were shifted between 0 and 1.2 nm, and 0.9 and 1.2 nm, respectively, using the standard MARTINI protocol[75]. An integration time step of 20 fs was applied. Covalent bonds were constrained to their equilibrium values using the LINCS

algorithm[79]. All simulations were run in the presence of standard MARTINI water particles[75], and ions added to an approximate concentration of 0.15 M NaCl.

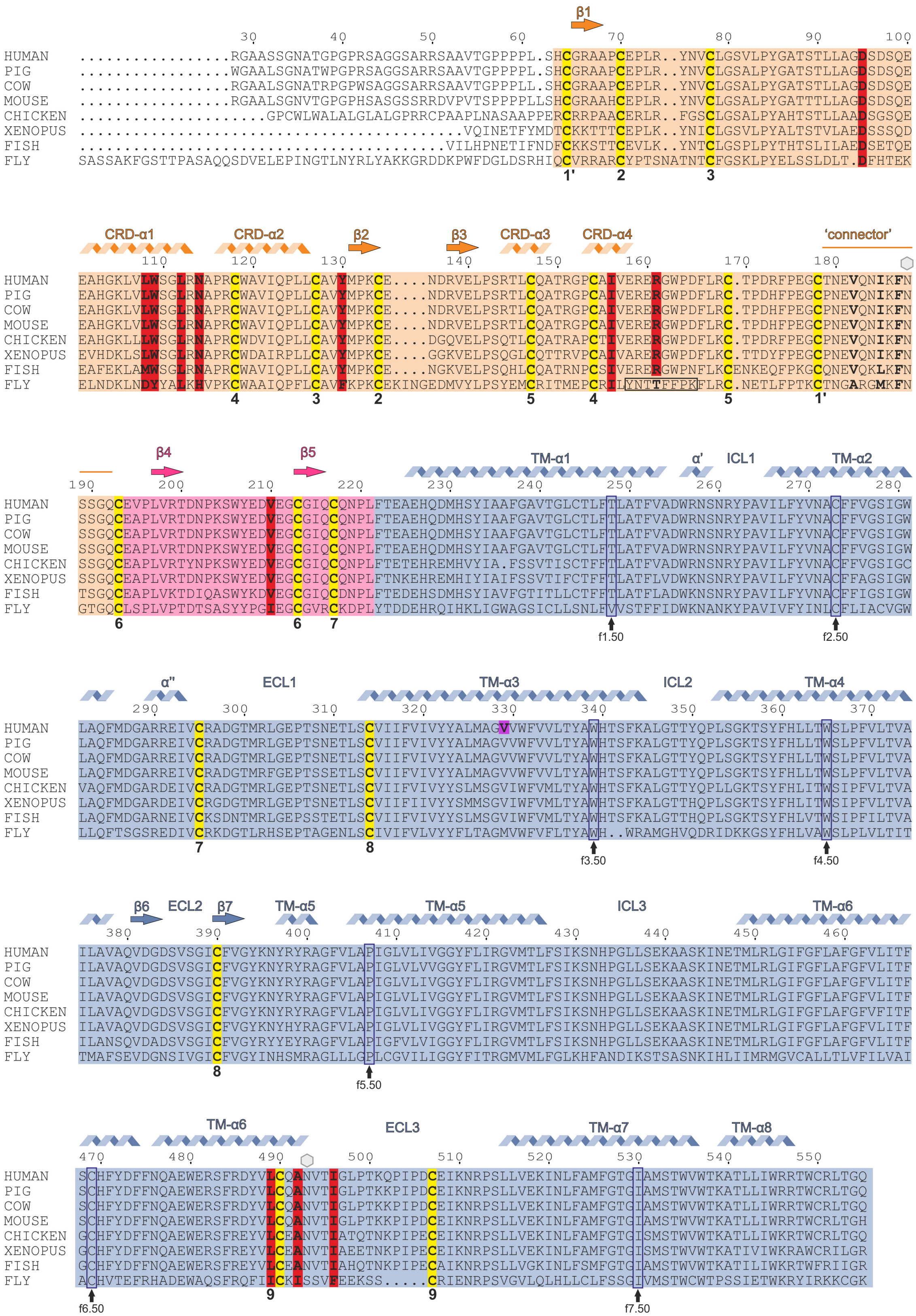
Atomistic simulations. Atomistic simulations were run using the GROMOS53a6 force field[80], and its extension to glycans[81]. The system was solvated using the SPC water model, and ions added to yield an electrically neutral system with a NaCl concentration of approximately 0.15 M. Systems contained approximately 140,000 atoms including 270 POPC molecules, ~41,000 water molecules, 149 sodium ions, and 154 chloride ions. Periodic boundary conditions were applied, with a simulation time step of 2 fs. Temperature was maintained at 310 K using a V-rescale thermostat[82] with a coupling constant of 0.1 ps, while pressure was controlled at 1 bar through coupling to a Parrinello–Rahman barostat[83] with a coupling constant of 1 ps. Particle Mesh Ewald (PME)[84] was applied to model long-range electrostatics. The LINCS algorithm was used to constrain covalent bond lengths[79]. The g_dist, g_rmsf and g_rms tools implemented in the GROMACS v4.6.3 software package[68] were applied to analyse the simulations, with VMD[85] and PyMOL[61] used for visualization. Cα r.m.s.d. calculations for the CRD were performed after a least-squares fit of the trajectory to Cα particles of the initial (crystal structure) CRD coordinates. Cα r.m.s.d. calculations for the TMD were performed after a least squares fit to Cα particles of the initial (crystal structure) transmembrane helix coordinates, excluding the inter-helix loop regions of the TMD from this calculation. DSSP matrices were produced using the do_dssp tool implemented in GROMACS[68].

SAXS experiments. For size-exclusion chromatography-coupled small-angle X-ray scattering (SEC–SAXS), amphipol-exchanged untreated and 20(S)-OHC-treated (~5 mM 20(S)-OHC) SMOΔC were loaded onto a 4.8-ml KW-403 column (Shodex), equilibrated in a no-detergent buffer, on an Agilent 1260 system (B21, Diamond Light Source). Approximately 45 μl of sample was injected at 9.6 (20(S)-OHC-treated) or 13 (apo) mg ml^{-1} using a flow rate of 160 μl per min. Chromatographic elution was directed into a specialized SAXS flow cell, with a 1.6 mm path length, held at 20 °C. SAXS measurements were made using a sample-to-detector distance of 4.09 m at a wavelength of 1 Å. SAXS images (frames) were collected as a continuous set of 3 s exposures across the elution peak. The corresponding buffer background frames for producing the background-subtracted SAXS curve was collected at greater than 1.5 column volumes.

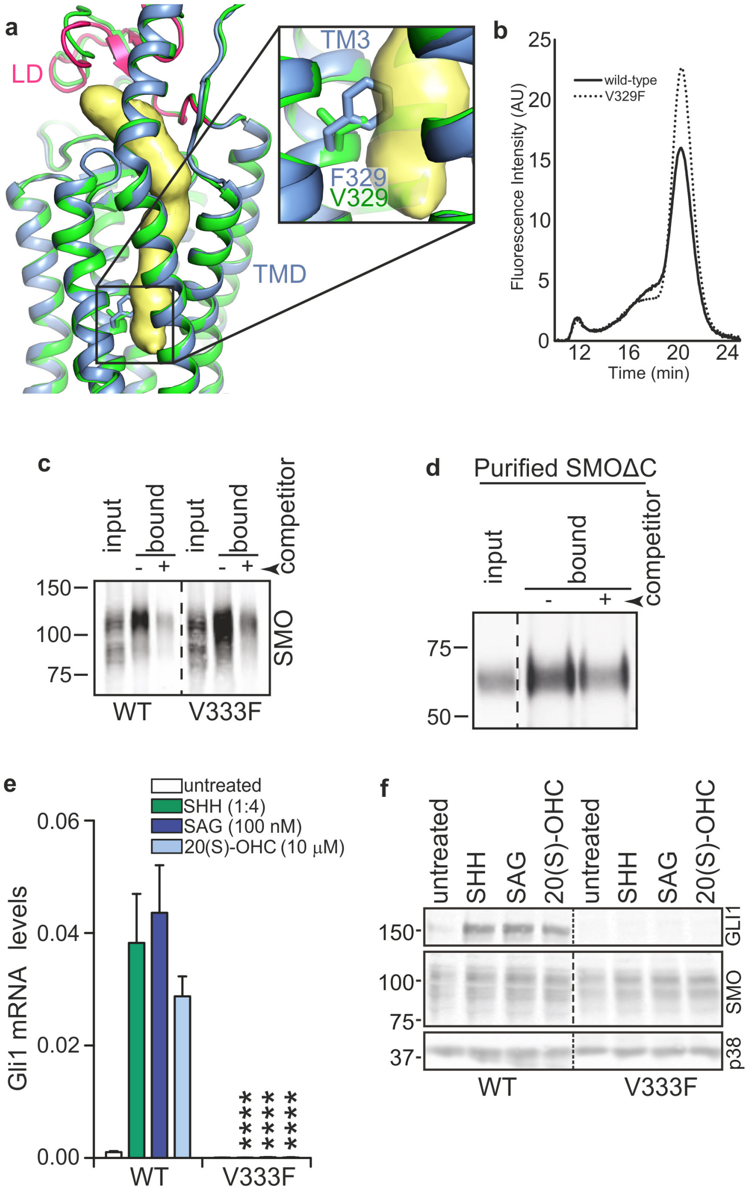
Images were corrected for variations in beam current, normalized for exposure time and processed into 1D scattering curves using in-house beamline software (GDA). Buffer subtractions and all other subsequent analysis were performed with the program ScÅtter (http://www.bioisis.net/scatter). Samples were checked for radiation damage by visual inspection of the Guinier region as a function of exposure time. Considerable differences between the SAXS curves ($q > 0.05$ Å^{-1}) of the treated and untreated samples imply major structural differences between the two states. Large differences between the two $P(r)$ distributions imply a significant structural change in the 20(S)-OHC state.

31. Varjosalo, M., Li, S. P. & Taipale, J. Divergence of hedgehog signal transduction mechanism between *Drosophila* and mammals. *Dev. Cell* **10,** 177–186 (2006).
32. Rohatgi, R., Milenkovic, L. & Scott, M. P. Patched1 regulates hedgehog signaling at the primary cilium. *Science* **317,** 372–376 (2007).
33. Humke, E. W., Dorn, K. V., Milenkovic, L., Scott, M. P. & Rohatgi, R. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes Dev.* **24,** 670–682 (2010). 10.1101/gad.1902910
34. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D Biol. Crystallogr*. **62,** 1243–1250 (2006).
35. Molday, R. S. & MacKenzie, D. Monoclonal antibodies to rhodopsin: characterization, cross-reactivity, and application as structural probes. *Biochemistry* **22,** 653–660 (1983).
36. Oprian, D. D., Molday, R. S., Kaufman, R. J. & Khorana, H. G. Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proc. Natl Acad. Sci. USA* **84,** 8874–8878 (1987).
37. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296,** 913–916 (2002).
38. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20,** 87–90 (2002).
39. Rohatgi, R., Milenkovic, L., Corcoran, R. B. & Scott, M. P. Hedgehog signal transduction by Smoothened: pharmacologic evidence for a 2-step activation process. *Proc. Natl Acad. Sci. USA* **106,** 3196–3201 (2009).
40. Miller, P. S. & Aricescu, A. R. Crystal structure of a human $GABA_A$ receptor. *Nature* **512,** 270–275 (2014).
41. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14,** 673–681 (2006).
42. Bailey, E. C., Milenkovic, L., Scott, M. P., Collawn, J. F. & Johnson, R. L. Several PATCHED1 missense mutations display activity in patched1-deficient fibroblasts. *J. Biol. Chem.* **277,** 33632–33640 (2002).
43. Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. USA* **90,** 8392–8396 (1993).
44. Ciepla, P. *et al.* New chemical probes targeting cholesterylation of Sonic Hedgehog in human cells and zebrafish. *Chem. Sci. (Camb.)* **5,** 4249–4259 (2014).
45. Chung, S.-K., Shim, J.-Y., Kang, M. G., Lee, K. W. & Kang, H. I. Studies of steroids as potential antifungal agent 2. Side chain modified cholesterols. *Korean J. Med. Chem.* **8,** 14–17 (1998).
46. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protocols* **4,** 706–731 (2009).
47. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43,** 186–190 (2010).
48. Kabsch, W. Xds. *Acta Crystallogr. D* **66,** 125–132 (2010).
49. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69,** 1204–1214 (2013).
50. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67,** 235–242 (2011).
51. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
52. Liu, W. *et al.* Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337,** 232–236 (2012).
53. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr*. **66,** 470–478 (2010).
54. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62,** 1002–1011 (2006).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
56. BUSTER v. 2.10.2 (Global Phasing Ltd., Cambridge, United Kingdom, 2011).
57. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
58. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
59. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98,** 10037–10041 (2001).
60. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179,** 125–142 (1984).
61. Schrodinger, L.L.C. The PyMOL Molecular Graphics System, Version 1.3r1 (2010).
62. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
63. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16,** 10881–10890 (1988).
64. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–W324 (2014).
65. Chovancova, E. *et al.* CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLOS Comput. Biol.* **8,** e1002708 (2012).
66. Tribet, C., Audebert, R. & Popot, J. L. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl Acad. Sci. USA* **93,** 15047–15050 (1996).
67. Gohon, Y. *et al.* Bacteriorhodopsin/amphipol complexes: structural and functional properties. *Biophys. J.* **94,** 3523–3537 (2008).
68. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4,** 435–447 (2008).
69. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).
70. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7,** 2284–2295 (2011).
71. de Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9,** 687–697 (2013).
72. Scott, K. A. *et al.* Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* **16,** 621–630 (2008).
73. Stansfeld, P. J. *et al.* MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **23,** 1350–1361 (2015).
74. Stansfeld, P. J. & Sansom, M. S. Molecular simulation approaches to membrane proteins. *Structure* **19,** 1562–1572 (2011).
75. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111,** 7812–7824 (2007).
76. Monticelli, L., Sorin, E. J., Tieleman, D. P., Pande, V. S. & Colombo, G. Molecular simulation of multistate peptide dynamics: a comparison between microsecond timescale sampling and multiple shorter trajectories. *J. Comput. Chem.* **29,** 1740–1752 (2008).
77. Periole, X., Cavalli, M., Marrink, S. J. & Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J. Chem. Theory Comput.* **5,** 2531–2543 (2009).

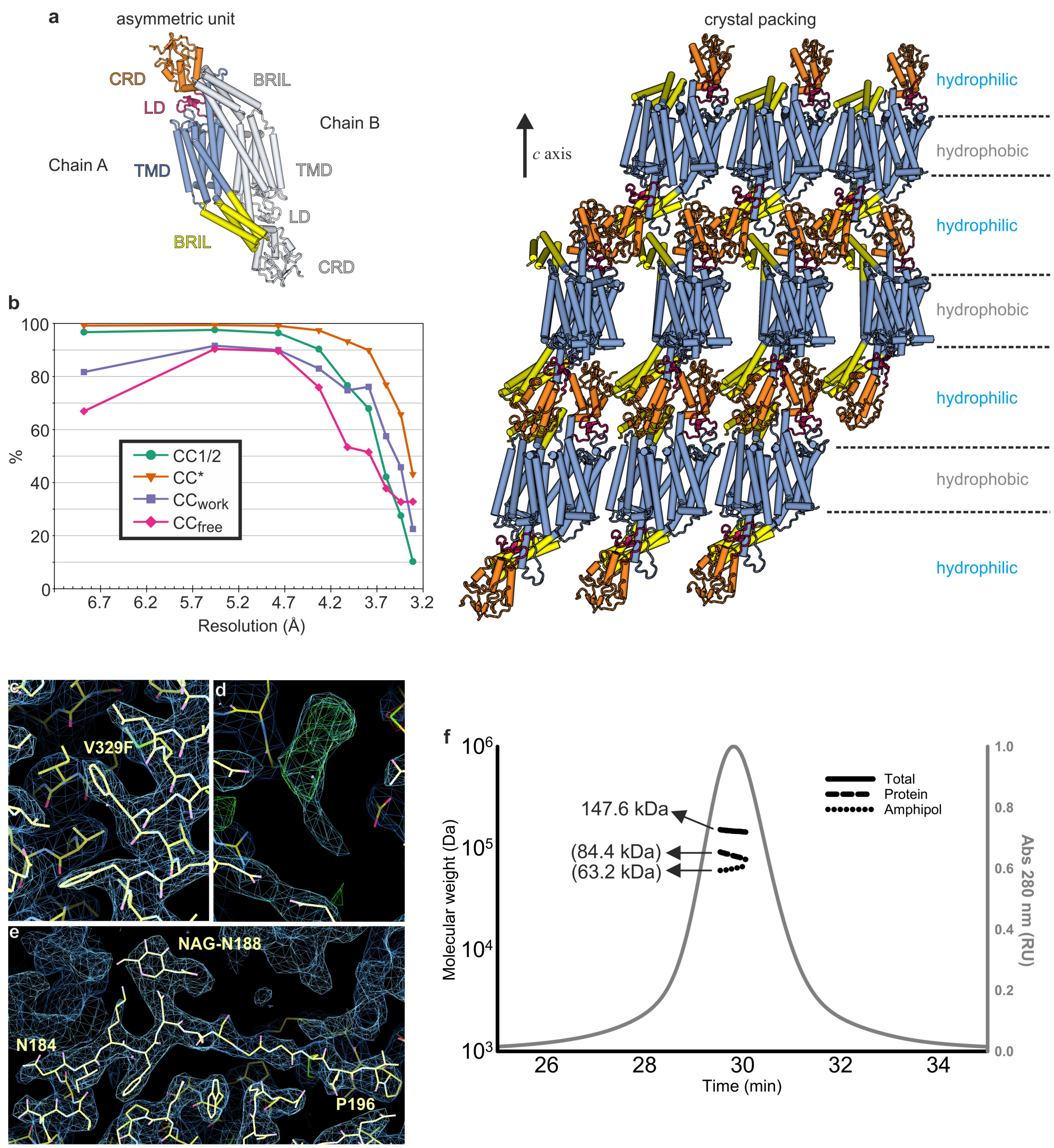
78. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).
79. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).
80. Oostenbrink, C., Villa, A., Mark, A. E. & van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25,** 1656–1676 (2004).
81. Pol-Fachin, L., Verli, H. & Lins, R. D. Extension and validation of the GROMOS 53A6(GLYC) parameter set for glycoproteins. *J. Comput. Chem.* **35,** 2087–2095 (2014).
82. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).
83. Parrinello, M. & Rahman, A. Polymorphic transitions in single-crystals — a new molecular-dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).
84. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).
85. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
86. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336,** 1030–1033 (2012).
87. Zoonens, M. & Popot, J. L. Amphipols for each season. *J. Membr. Biol.* **247,** 759–796 (2014).



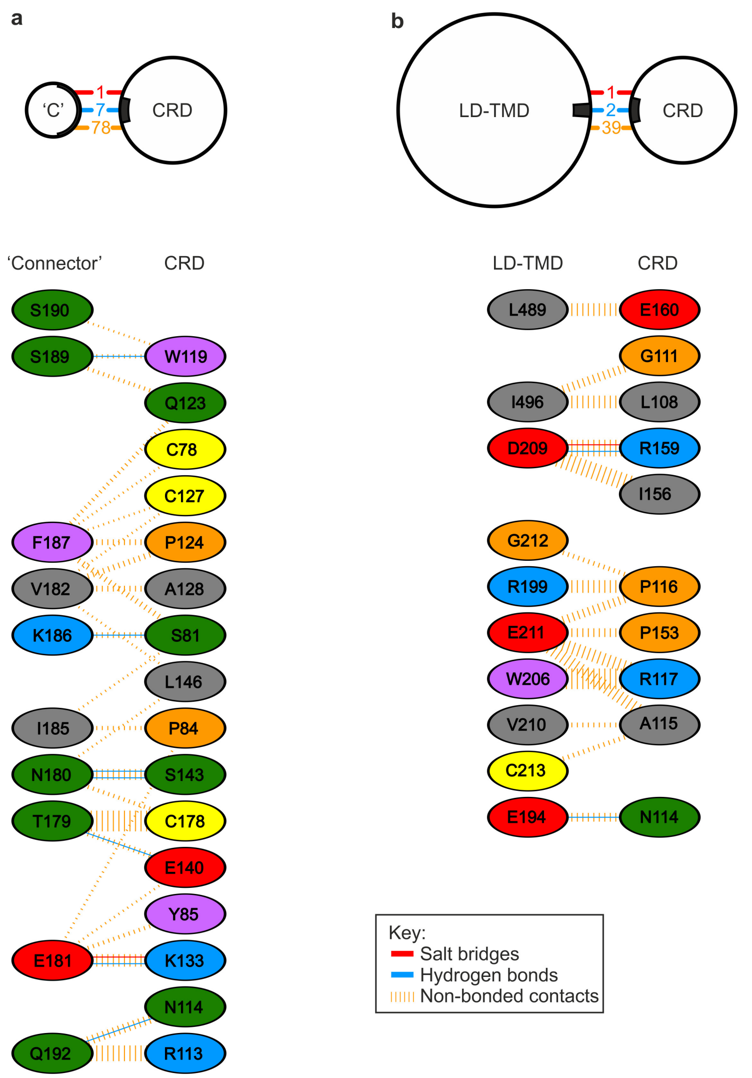
Extended Data Figure 1 | Sequence alignment of SMO orthologues. Numbering corresponds to that of human SMO. Secondary structure assignments are displayed above the alignment and colour-coded as in Fig. 1. Black arrows and numbers (fX.50) below alignment show class F Ballesteros–Weinstein nomenclature for GPCR helices[16]. Residues interacting with cholesterol are highlighted in red. Disulfide bridges are highlighted in yellow and numbered. *N*-linked glycosylation sites are depicted by a hexagon. The position of the Val329Phe point mutation is highlighted in purple.



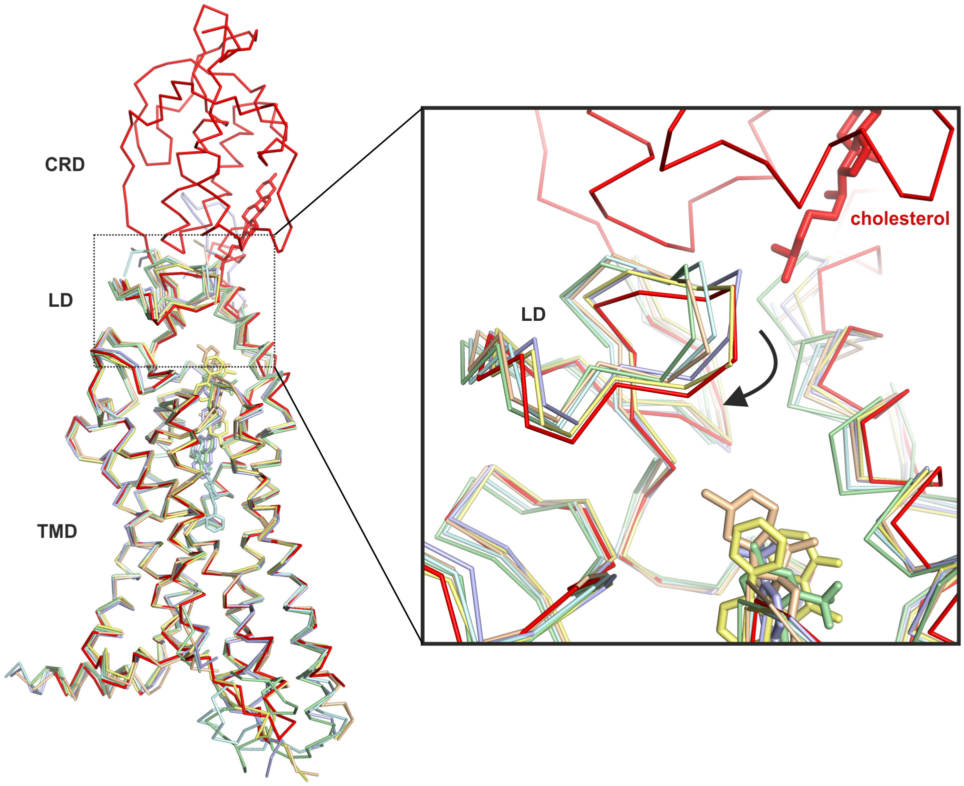
Extended Data Figure 2 | Characterization of the SMO Val329Phe mutation. a, Superposition of SMOΔC structure (blue) with the SMO-SANT-1 complex structure, which lacks the CRD (green, PDB 4N4W (ref. 16)), showing the TMD ligand-binding pocket as a yellow surface. Inset shows Val329, mutated to Phe in our structure. **b**, SEC analysis of fluorescently labelled SMOΔC showing difference in expression levels of wild-type and Val329Phe variant (main protein peak ~20 min). **c**, 20(S)-OHC beads can bind both mouse wild-type SMO and Val333Phe (mouse Val333 corresponds to human Val329). Immunoblots, using an anti-SMO antibody directed against the ICD, were used to measure SMO captured on 20(S)-OHC beads. Adding 50 μM free 20(S)-OHC as a competitor reduced binding. **d**, Purified human SMOΔC (the crystallization construct) binds to 20(S)-OHC beads. **e**, *Smo*$^{-/-}$ mouse fibroblasts stably expressing SMO-WT or SMO-Val333Phe were exposed to SHH, SAG or 20(S)-OHC. Levels of endogenous *Gli1* mRNA (mean arbitrary units ± s.d., $n = 4$), measured by qRT–PCR, were used as a metric of Hh pathway activity because *Gli1* is a direct Hh target. Asterisks indicate statistical significance (****$P \leq 0.0001$) based on one-way ANOVA for the difference in *Gli1* mRNA levels between identically treated SMO-WT and SMO-Val333Phe cells. **f**, Immunoblot shows SMO and GLI1 protein levels in these stable cell lines, with p38 as loading control. Each experiment was replicated ≥2 times with similar results.



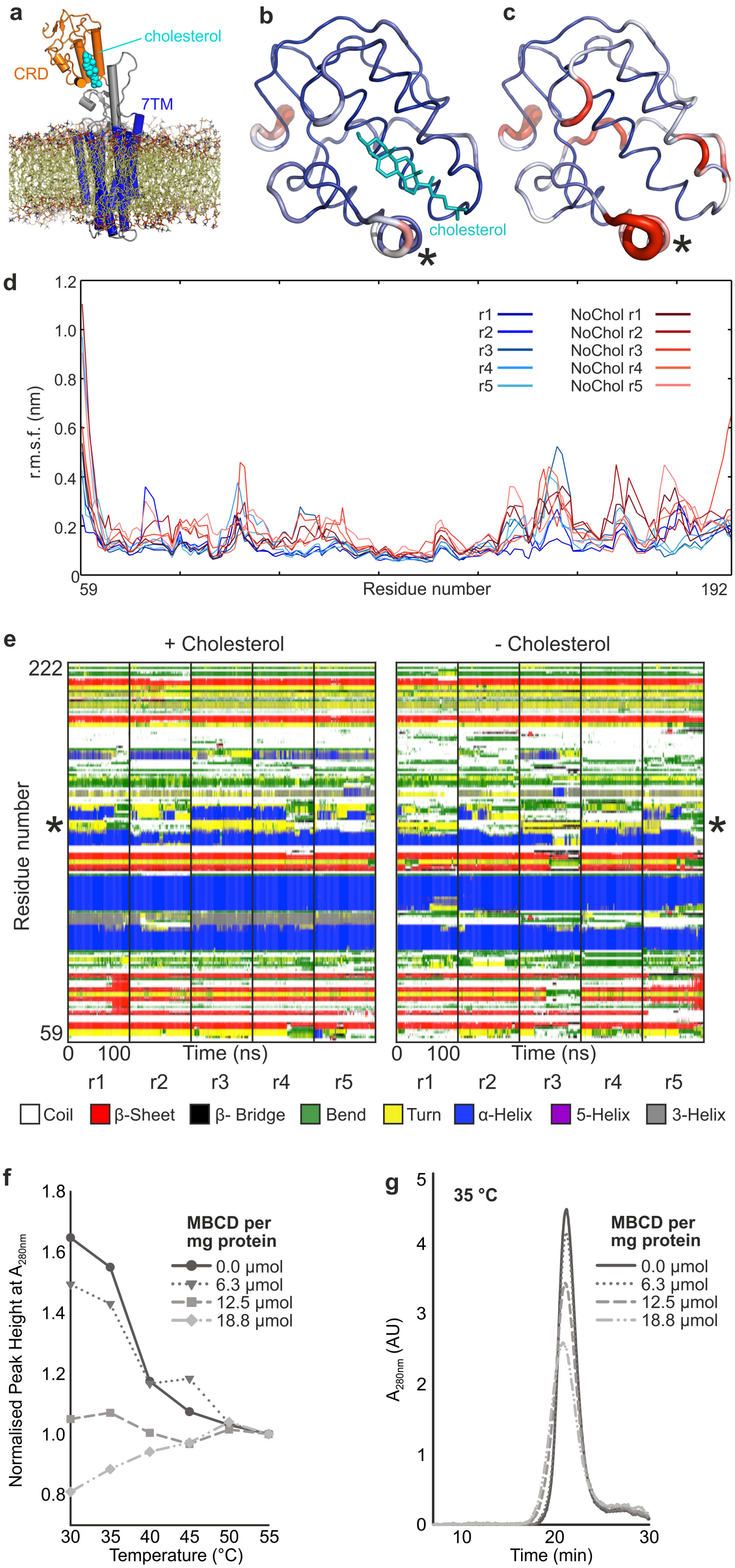
Extended Data Figure 3 | Crystallization, structure solution and oligomeric state of SMOΔC. a, SMOΔC crystal packing. Asymmetric unit consists of two antiparallel SMOΔC chains. Chain A coloured as in Fig. 1 with BRIL fusion in yellow; Chain B in grey. LCP crystal packing with alternating hydrophobic and hydrophilic layers perpendicular to the *c* axis. Molecules coloured as for Chain A. **b**, Pearson correlation coefficient (CC) analysis[86] used to relate data quality with model quality. A CC_{work} and CC_{free} smaller than CC* indicates that the model does not account for all of the signal in the data (and is therefore not overfit). **c–e**, SigmaA-weighted $2F_o–F_c$ electron density maps of final refinement at 1.0σ contour level. **c**, Val329Phe mutation. **d**, Extra density within TMD ligand-binding pocket ($F_o–F_c$ maps shown at contour level of $+3\sigma$ (green) and -3σ (red)) (This density could not be confidently assigned, probably because of low occupancy within the crystal.). **e**, 'Connector' region linking the CRD and linker domain, with Asn188 and linked *N*-acetyl glycosamine moiety. **f**, SEC–MALS analysis of amphipol-solubilized SMOΔC. Molar masses (M_W, black lines) and 280 nm absorption (grey line) plotted against elution time. M_W derived from protein-conjugate analysis indicated in parentheses. For clarity, graphs of M_W are shown only around main absorption peak. Theoretical M_W of SMOΔC based on sequence is 71 kDa with the extra mass observed in MALS (~13 kDa) probably due to three *N*-linked glycosylation sites. Expected mass of protein-bound amphipol (A8–35) was previously determined to be 40–75 kDa (ref. 87), in agreement with our data. This analysis suggests that SMOΔC is a monomer under our purification conditions.



Extended Data Figure 4 | Interfaces of the SMO CRD. a, **b**, Detailed interactions of the CRD with the connector region (**a**) and with the linker domain–TMD segment (**b**). The number of interactions are indicated in the top panel and coloured as indicated in the key box. For non-bonded contacts, the width of the striped line is proportional to the number of atomic contacts. Residue colouring is according to amino acid: blue, positive (H,K,R); red, negative (D,E); green, neutral (S,T,N,Q); grey, aliphatic (A,V,L,I,M); purple, aromatic (F,Y,W); orange, proline (P) or glycine (G); yellow, cysteine (C). The figure is adopted from the PDBSUM server (http://www.ebi.ac.uk/pdbsum/).

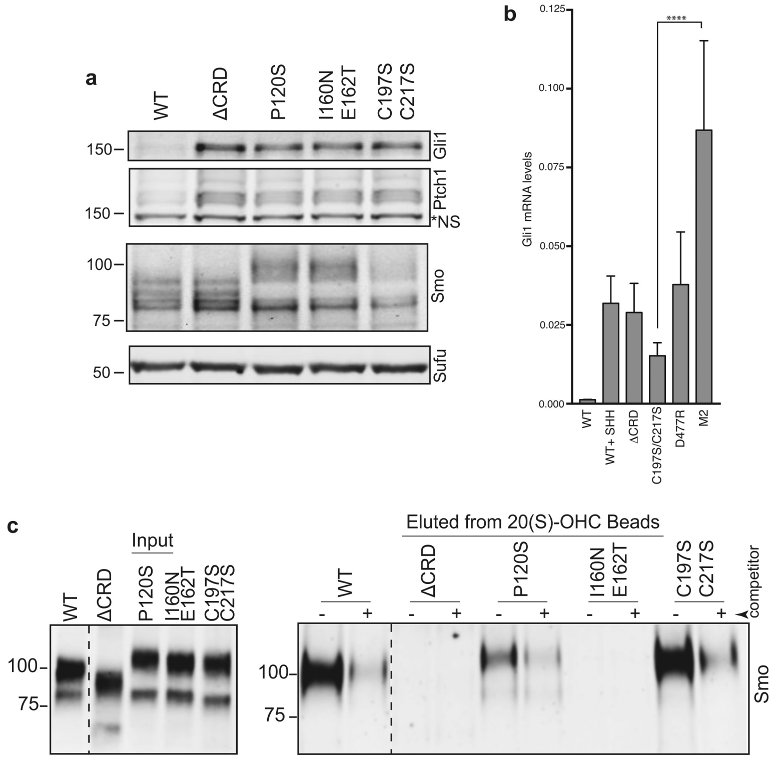


Extended Data Figure 5 | Comparison of SMOΔC with previously determined SMO TMD structures. Superposition of the SMOΔC structure with SMO TMD structures. Structural alignment was performed using the 7TM bundle as template (not including the linker domain or TMD helix 8). SMOΔC (red), SMO TMD complexed with cyclopamine (light orange, PDB 4O9R, r.m.s.d. 0.598 Å for 243 equivalent Cα positions), antaXV (light blue, PDB 4QIM, r.m.s.d. 0.515 Å for 233 equivalent Cα positions), SANT1 (pale cyan, PDB 4N4W, r.m.s.d. 0.483 Å for 240 equivalent Cα positions), LY2940680 (pale green, PDB 4JKV, r.m.s.d. 0.493 Å for 230 equivalent Cα positions), SAG1.5 (pale yellow, PDB 4QIN, r.m.s.d. 0.623 Å for 262 equivalent Cα positions). The box shows a close-up view of the linker domain region revealing a structural rearrangement in the SMOΔC structure compared to the previously determined SMO TMD structures lacking the native extracellular domain.

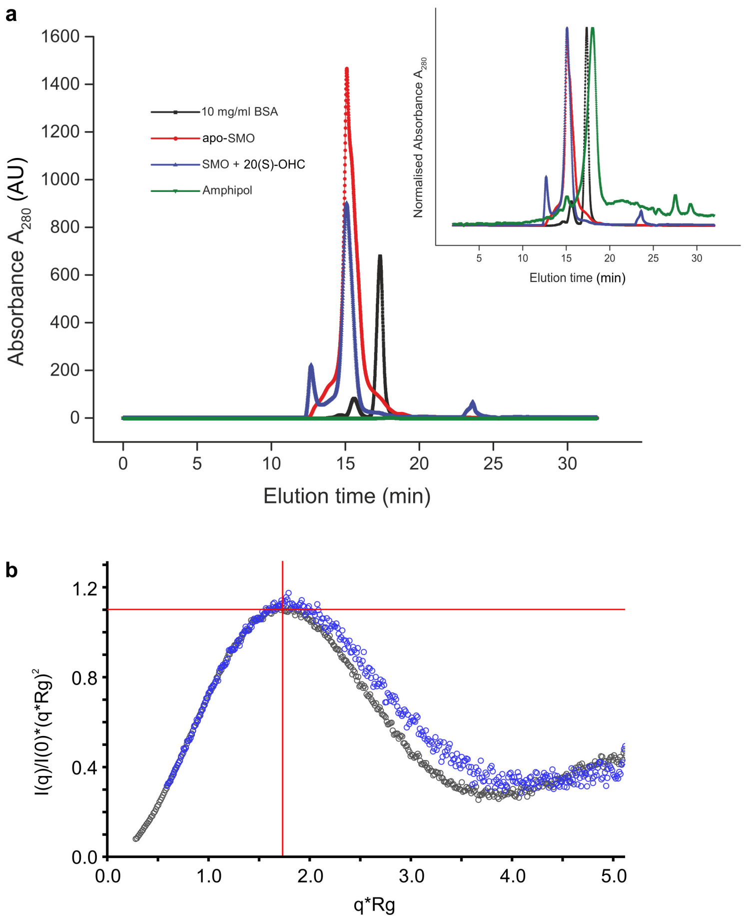


Extended Data Figure 6 | See next page for caption.

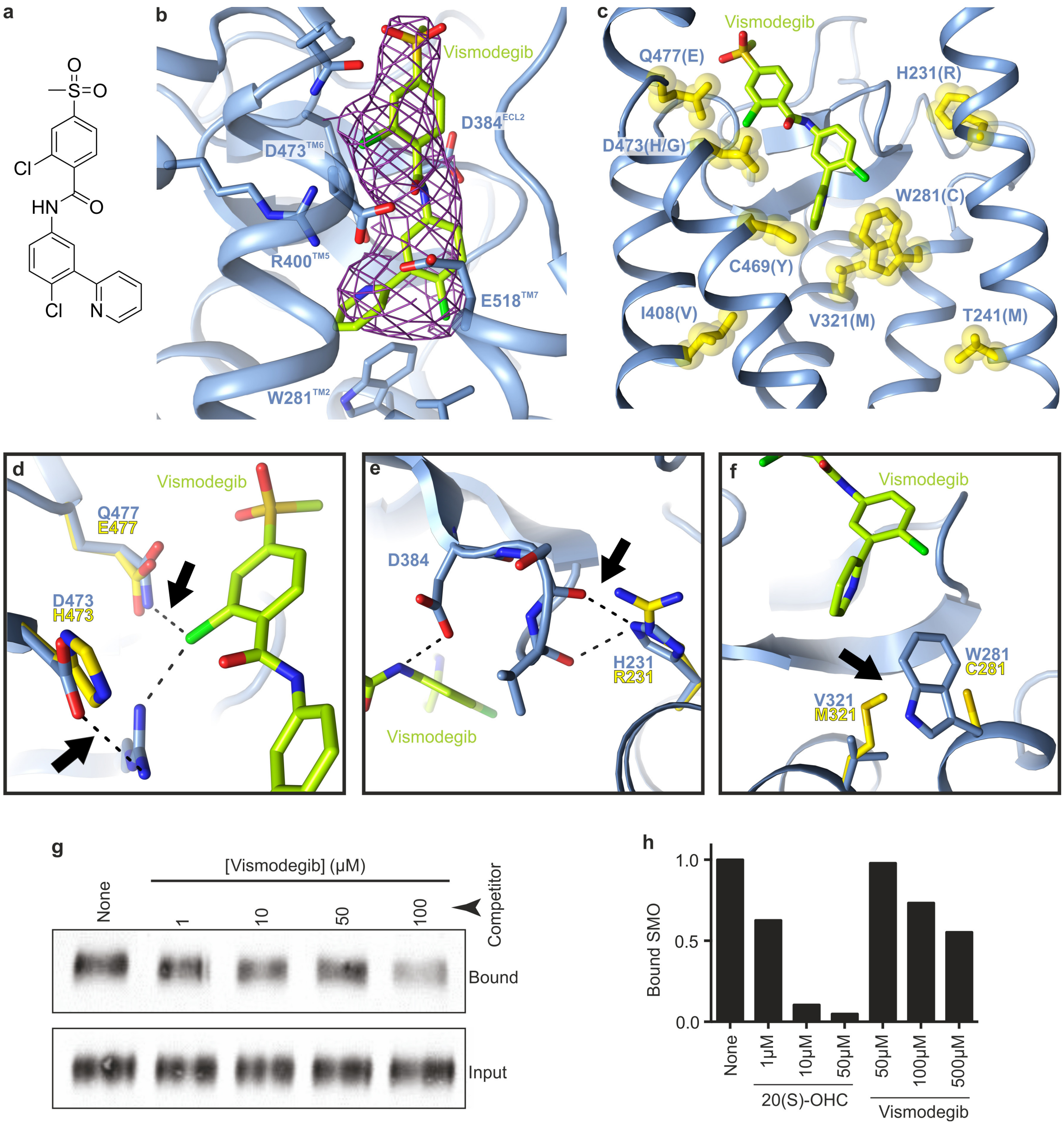
Extended Data Figure 6 | Cholesterol stabilizes SMO. a–**e**, MD simulations of SMO in a lipid bilayer. **a**, SMO embedded in a lipid bilayer with the CRD in orange, the seven-pass transmembrane region excluding intra- and extracellular loops (7TM) in blue and cholesterol in cyan. **b**–**d**, Relative r.m.s. fluctuations of the Cα atoms over the course of 5×100 ns of atomistic MD simulation in the presence and absence of cholesterol. The structures in **b** and **c** are shown as putty representations coloured from high conformational stability (that is, low r.m.s. fluctuations; blue/thin) to low stability (that is, high r.m.s. fluctuations; red/thick). **e**, Secondary structure DSSP matrices for each of the simulations. The asterisks in **b**, **c** and **e** all mark the helix spanning residues 155–160, which is destabilized in the absence of bound cholesterol. **f**, **g**, Thermostability of purified SMOΔC. See Supplementary Discussion for details. **f**, Compiled peak heights from thermostability SEC analysis of purified SMOΔC after treatment with different MBCD concentrations. **g**, Example of raw SEC data used for the analysis in **f**. Samples were incubated at 35 °C for 1 h before loading onto the SEC column.



Extended Data Figure 7 | Effect of domain interface mutations on expression levels and 20(S)-OHC binding. a, Protein levels of SMO and also of PTCH1 and GLI1 (each of which is encoded by a direct Hh target gene) measured by immunoblot from *Smo*$^{-/-}$ mouse fibroblasts stably expressing one of five SMO variants: wild-type SMO (WT); SMO lacking the entire CRD (ΔCRD); SMO with two mutations (Pro120Ser or Ile160Asn/Glu162Thr) that introduce glycosylation sites in the linker domain–CRD interface; and SMO lacking a conserved disulfide bond (Cys197Ser/Cys217Ser, marked 6 in Fig. 4c) in the linker domain. Elevated levels of GLI1 and PTCH1 reflect high constitutive signalling activity of each mutant. NS, a non-specific band detected by anti-PTCH1 antibody; SUFU, loading control. Different patterns seen in SMO panel are caused by different numbers of *N*-linked glycosylation sites. **b**, *Gli1* mRNA levels (mean arbitrary units ± s.d., $n = 3$) were used to assess Hh signalling activity in *Smo*$^{-/-}$ cells stably expressing the indicated mouse SMO variants. One-way ANOVA was used to assess statistical significance (****$P \leq 0.0001$). D477R and M2 (Trp539Leu) are two previously described mutations in the TMD that increase constitutive signalling. **c**, Oxysterol-binding capacity of each SMO variant was determined (right blot) by its ability to bind to 20(S)-OHC beads in the absence or presence of 50 μM free 20(S)-OHC. Inputs for each binding reaction are shown on the left. Each experiment was repeated 2 or more times with similar results.



Extended Data Figure 8 | SAXS analysis of SMOΔC. **a**, Overlay of size-exclusion chromatograms monitored at 280 nm (A_{280}) collected during SAXS measurements for *apo*-SMOΔC (red), (+)20(S)-OHC SMOΔC (blue), amphipol (green) and BSA standard (black). Amphipol and BSA were injected at 10 mg ml^{-1}. Inset shows curves normalized to peak height. BSA was used as a reference with a radius-of-hydration of 3.7 nm. Absorbance of the free amphipol is negligible and elutes ~5 min after the amphipol-stabilized SMOΔC samples. **b**, Dimensionless Kratky plot of *apo*- and (+)20(S)-OHC-loaded SMOΔC SAXS data. Cross-hairs denote the Guinier–Kratky point ($\sqrt{3}$, 1.1), the peak position for an ideal, globular particle. The slower decay of the transformed scattering intensities for (+)20(S)-OHC (blue) indicate a comparably less spherical particle.



Extended Data Figure 9 | Crystal structure of the SMOΔC–vismodegib complex and structural analysis of mutations found in vismodegib-resistant cancers. a, Chemical structure of vismodegib. **b**, Close-up view of vismodegib-binding site. Colour-coding follows Fig. 5b. Composite omit map calculated with PHENIX at 1.0σ shown as magenta chicken-wire. **c**, Mapping of residues that are mutated in vismodegib-resistant tumours (yellow highlights). Brackets indicate mutant residues. **d**–**f**, Close-up views of selected interactions. Native residues in blue and mutated residues in yellow. Arrows indicate position of potential clashes that could disrupt vismodegib binding. **d**, Gln477/Asp473 hot spot. The Gln477Glu mutation leads to a loss of the potential hydrogen bond of the glutamine sidechain to the chloride of the vismodegib chlorophenyl-methylsulfone moiety. The Asp473His mutation potentially destabilizes the hydrogen-bonding network around Arg400 that coordinates the vismodegib chlorophenyl-methylsulfone moiety. **e**, The imidazole ring of His231 is within hydrogen-bonding distance of two carbonyl main-chain atoms of residues Ser385 and Val386 located on a loop coordinating the interaction of Asp384 with vismodegib's amide linker. **f**, Trp281 forms a key hydrophobic interaction with the vismodegib pyrimidine ring that is deeply buried in the SMO helical bundle core. Mutation to cysteine would significantly destabilize this interaction while mutation of nearby Val321 to the bulkier methionine would probably result in a rearrangement of the Trp281 side chain. **g**, **h**, SMOΔC captured on cholesterol beads in the presence of increasing concentrations of free vismodegib or 20(S)-OHC (**h**). Results from one of two independent pull-down experiments are shown.

Extended Data Table 1 | Crystallographic data collection and refinement statistics

	apo-SMOΔC	*vismo*-SMOΔC
Data collection		
Beamline	DIAMOND-I24	DIAMOND-I24
Wavelength (Å)	0.9686	0.9686
No. crystals	9	2
Resolution (Å)	61.0-3.2 (3.41-3.20)	68.4-3.3 (3.39-3.30)
Space group	C2	P1
Cell dimensions		
a, b, c (Å)	122.9, 63.0, 208.6	52.8, 71.3, 107.0
α, β, γ (°)	90.0, 96.6, 90.0	91.5, 98.2, 105.9
Unique reflections	25214 (3815)	21097 (2380)
Completeness (%)	93.8 (78.8)	94.7 (85.8)
R_{merge} (%)	30.5 (>100)	15.2 (>100)
R_{pim} (%)	13.8 (>100)	9.9 (>100)
$CC_{1/2}$	0.980 (0.237)	0.953 (0.357)
CC*	0.995 (0.580)	0.988 (0.460)
CC_{work}	0.923 (0.354)	0.703 (0.340)
CC_{free}	0.941 (0.333)	0.941 (0.315)
I/σI	4.3 (0.6)	3.7 (0.6)
Multiplicity	5.8 (5.1)	1.8 (1.7)
Refinement		
Resolution (Å)	61.0-3.2	60.0-3.3
No. reflections (test set)	25151 (1302)	20588 (1013)
R_{work}/R_{free} (%)	23.2/26.4	24.5/25.7
No. atoms		
Protein	9207	8988
Ligand/Ion	58	119
B factors (Å^2)		
Protein	137	124
Ligand/Ion	121	119
r.m.s.d. bonds (Å)	0.008	0.008
r.m.s.d. angles (°)	0.90	0.87
Ramachandran (%)		
Favoured	96.6	96.8
Allowed	3.2	2.9
Disallowed	0.2	0.3

Numbers in parentheses refer to the highest resolution shell. R_{free}, R-factor against 5% of the data; r.m.s.d., root mean square deviation from ideal geometry; R_{pim}, precision indicating merging R-factor; $CC_{1/2}$, correlation coefficient between random half data sets; CC*, analytical estimate of CC_{true} based on $CC_{1/2}$; CC_{work}, standard correlation of the experimental intensities with the intensities calculated from the refined model; CC_{free}, cross-validated correlation of the experimental intensities with the intensities calculated from the refined model. Ramachandran statistics were calculated using MolProbity.

LETTER

doi:10.1038/nature18292

Suppression of star formation in dwarf galaxies by photoelectric grain heating feedback

John C. Forbes[1], Mark R. Krumholz[1,2], Nathan J. Goldbaum[3] & Avishai Dekel[4]

Photoelectric heating—heating of dust grains by far-ultraviolet photons—has long been recognized as the primary source of heating for the neutral interstellar medium[1]. Simulations of spiral galaxies[2] have shown some indication that photoelectric heating could suppress star formation; however, simulations that include photoelectric heating have typically shown that it has little effect on the rate of star formation in either spiral galaxies[3,4] or dwarf galaxies[5], which suggests that supernovae are responsible for setting the gas depletion time in galaxies[6–8]. This result is in contrast with recent work[9–13] indicating that a star formation law that depends on galaxy metallicity—as is expected with photoelectric heating, but not with supernovae—reproduces the present-day galaxy population better than does a metallicity-independent one. Here we report a series of simulations of dwarf galaxies, the class of galaxy in which the effects of both photoelectric heating and supernovae are expected to be strongest. We simultaneously include space- and time-dependent photoelectric heating in our simulations, and we resolve the energy-conserving phase of every supernova blast wave, which allows us to directly measure the relative importance of feedback by supernovae and photoelectric heating in suppressing star formation. We find that supernovae are unable to account for the observed[14] large gas depletion times in dwarf galaxies. Instead, photoelectric heating is the dominant means by which dwarf galaxies regulate their star formation rate at any given time, suppressing the rate by more than an order of magnitude relative to simulations with only supernovae.

To investigate whether the depletion times in dwarf galaxies, which are longer than those for Milky-Way-like galaxies by more than an order of magnitude[14,15], are set by the momentum injection from supernovae or by photoelectric heating, we perform a series of high-resolution hydrodynamic simulations using the Enzo adaptive mesh refinement code[16]. We additionally include a prescription for supernova and pre-supernova stellar feedback, and a method for self-consistent, spatially dependent photoelectric heating (see Methods). We use two sets of initial conditions. Both correspond to isolated dwarf galaxies with an initially laminar gas disk with a gas surface density profile that falls off exponentially, a stationary hot halo, and collisionless particles representing stars and dark matter. The galaxies have a dark matter halo mass of 10^{10} solar masses ($M_\odot$), a stellar mass of $10^7 M_\odot$ and an observationally motivated[17] cold gas mass of $10^8 M_\odot$ Galaxies in this mass range are comfortably above the limit at which star formation can be quenched by the cosmological ultraviolet background[18], but small enough that the effects of heating by far-ultraviolet photons and supernova feedback[19] are plausibly extreme. The initial conditions differ in the exponential scale-length chosen for the gas. One set uses 5 kpc, which is designed to mimic recently discovered, nearly starless galaxies[20], and which is towards the high end of the range observed for field dwarf galaxies[14]; the other set uses 1 kpc, which is towards the low end of the observed field dwarf range. For a galaxy with an H I mass of $10^8 M_\odot$, and assuming an exponential H I profile, the observed relation[21] between H I mass and H I size suggests an H I scale-length of about 1.9 kpc.

To understand how supernovae and photoelectric heating each contribute to the evolution of these galaxies, we perform a straightforward numerical experiment. We run a fiducial simulation, starting from initial conditions with a 5-kpc gas scale-length, that includes supernovae and photoelectric heating, and simulations for which each of these effects is turned off in turn. We refer to these as the 'SN + PE', 'PE only', 'SN only' and the 'no feedback' simulations. We also run the 'SN + PE', 'PE only' and 'no feedback' cases for initial conditions with a 1-kpc scale-length. The simulations with supernovae also include pre-supernova stellar feedback from winds and H II regions. For the 5-kpc case, we re-run the simulations of the four feedback models at three maximum spatial resolutions: 10 pc, 5 pc and 2.5 pc. These resolutions are high enough, and the typical densities in which supernovae explode in these simulations are low enough, that our simulations do not suffer from the overcooling problem[22], whereby poorly resolved simulations overestimate the rate at which supernova-heated gas cools (see Extended Data Fig. 1). We focus first on the 10-pc-resolution simulations with the 5-kpc initial conditions, because we ran these for the longest time. In Methods, we compare all our simulations to the higher

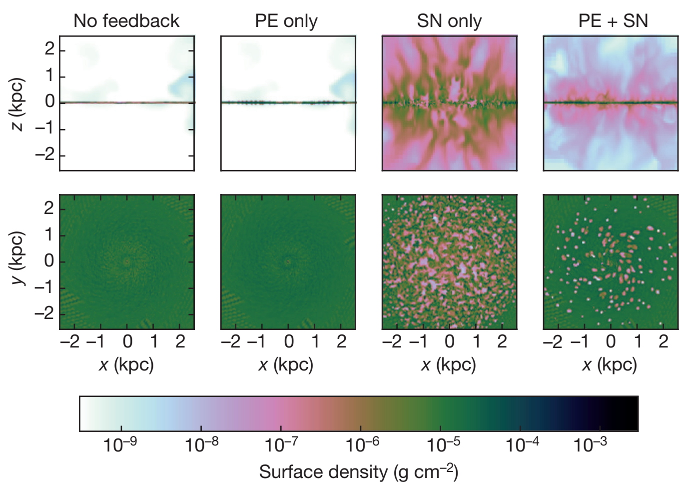


Figure 1 | The morphology of the gas. For each of the simulations, the surface density after 90 Myr of evolution, starting from initial conditions with a 5-kpc scale-length, is integrated between ±200 pc in the *y* dimension (top panels) and in the *z* dimension (bottom panels). The morphology of the disk is essentially determined by the presence of supernovae, despite the fact that the 'PE only' and 'PE + SN' runs have nearly identical star formation rates (Fig. 2). The star formation rate in the 'SN only' simulation is an order of magnitude higher than that in the 'PE + SN' simulation, which is why the outflow and the disruption of the cold disk are more striking in the former.

[1]Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA. [2]Research School of Astronomy and Astrophysics, Australian National University, Canberra, Australian Capital Territory 2611, Australia. [3]National Center for Supercomputing Applications, University of Illinois, 1205 West Clark Street, Urbana-Champaign, Illinois 61820, USA. [4]Center for Astrophysics and Planetary Sciences, Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel.

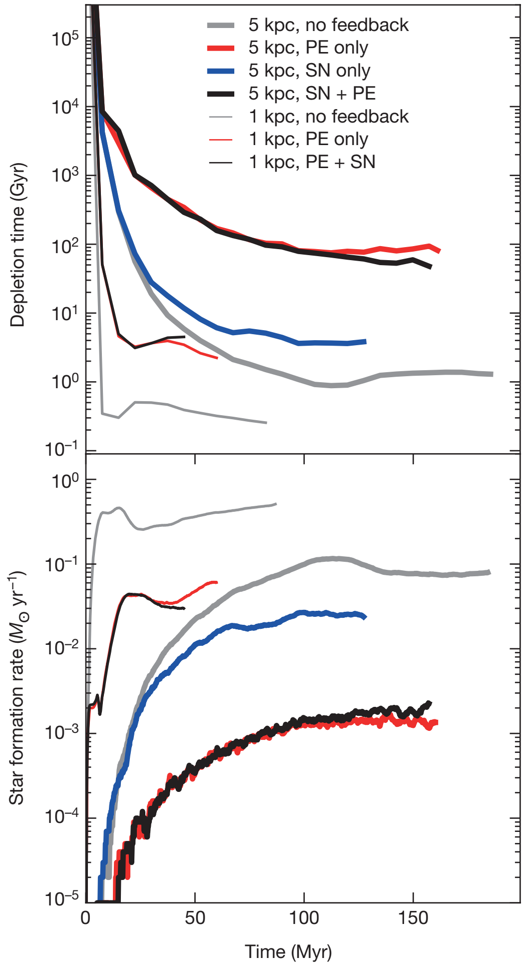


Figure 2 | The star formation rates of the simulations. The long depletion times frequently observed in dwarfs (see Extended Data Fig. 2) are reproduced when photoelectric heating and supernova feedback are both included ('PE + SN', black lines). Turning off supernova feedback has almost no effect (compare the red and black lines), which implies that photoelectric heating alone is responsible for the long depletion times in the simulations. More compact galaxies (1-kpc initial scale-length; thin lines) have higher star formation rates and shorter depletion times than do less compact galaxies (5-kpc initial scale-length; thick lines), but, as in the extended case, photoelectric heating remains the dominant effect.

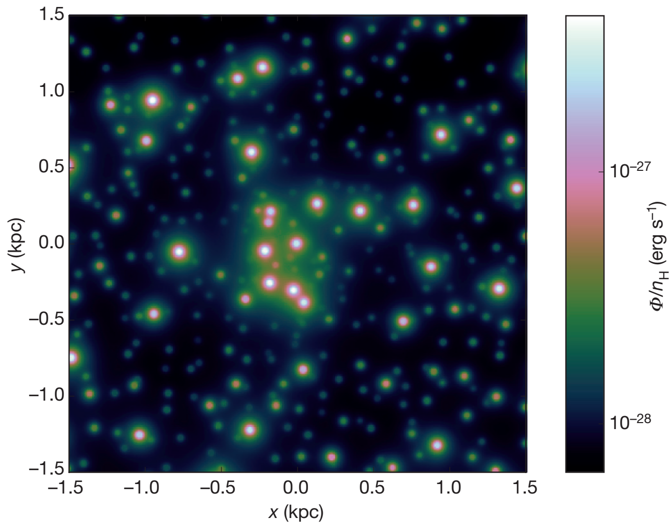


Figure 3 | The photoelectric heating rate. Every young star particle in the simulation produces far-ultraviolet radiation. The flux from all of these stars is summed in each cell (see Methods), yielding the volumetric heating rate Φ/n_H shown here. Given the low surface density of young stars, this distribution is highly inhomogeneous. The low metallicity and small size of the galaxy means that the mean free path of the far-ultraviolet photons is large, so the flux from each star and, hence, the heating rate falls off as r^{-2} from each source (where r is the radial coordinate).

resolution runs to evaluate the level of convergence of our results. We compare the 5-kpc and 1-kpc initial conditions below.

We find that all of the simulations follow a similar initial transient behaviour. The gas disk cools from its centre outwards, causing the disk to collapse vertically. Stars form first in the centre, then further and further out. The central region of the galaxy after 90 Myr is shown (with the aid of the yt package[23]) in Fig. 1, for each of the feedback models. In terms of large-scale morphology, the supernovae have the most noticeable effect, driving large outflows with mass loading factors of the order of 10. Photoelectric heating slightly alters the global structure of the gas, but the two simulations without supernovae look quite similar.

Figure 2 shows the star formation rates and depletion times as a function of time for each of the four feedback models. Each simulated galaxy experiences an initial transient period as the gas collapses from its initial state; however, the subsequent instantaneous star formation rates and depletion times of the galaxies are markedly different depending on whether photoelectric heating was included or not. By contrast, two simulations that differ only in their inclusion of supernova feedback result in similar depletion times. This immediately shows that photoelectric heating, not feedback from supernovae, is primarily responsible for the long depletion times observed in dwarf galaxies. Supernovae, or even a lack of photoelectric heating, can result in low star formation rates in the long run by rapidly ejecting gas or by locking gas into stellar remnants, respectively. This can be seen in the central kiloparsecs of the simulations without photoelectric heating: the star formation rate decreases in this region, but the depletion time is relatively unaffected. Only the simulations including photoelectric heating produce depletion times in reasonable agreement with the large values frequently observed in dwarf galaxies, as illustrated in Extended Data Fig. 2.

We also check the dependence of our results on the gas scale-length. In Fig. 2, the thin lines show the results for our simulations with a 1-kpc gas scale-length. We find that the more compact galaxies have higher star formation rates and shorter depletion times than do less compact galaxies, which explains the wide range of depletion times visible in the observational data. However, even for the 1-kpc simulations we find that the star formation rate is virtually unaffected when supernovae are disabled while keeping photoelectric heating active. By contrast, disabling photoelectric feedback causes the depletion time to decrease by an order of magnitude, to values that are inconsistent with the observed sample. This indicates that photoelectric heating, not supernovae, regulates star formation over a wide range of gas surface density in dwarf galaxies, and not just in the potentially extreme[20,24], nearly starless dwarf galaxies.

The means by which photoelectric heating suppresses star formation in our simulations is simple and intuitive. When a new star cluster is formed in the simulation, its most massive stars will emit photons with energies of 8–13.6 eV. These photons dominate the heating rate, owing to the grain photoelectric effect, because they have energies high enough to liberate electrons from dust grains, but low enough not to be absorbed by intervening neutral hydrogen (Fig. 3). The heating rate in the vicinity of a newly formed star increases the equilibrium temperature of the gas at fixed pressure or density. This in turn increases the Jeans mass of the gas and makes star formation more difficult (Fig. 4).

We conclude that the physics responsible for setting the instantaneous star formation law in dwarf galaxies—that is, the depletion time—is nearly independent of the physics determining the properties of the outflows. Ultimately young stars are responsible for photoelectric

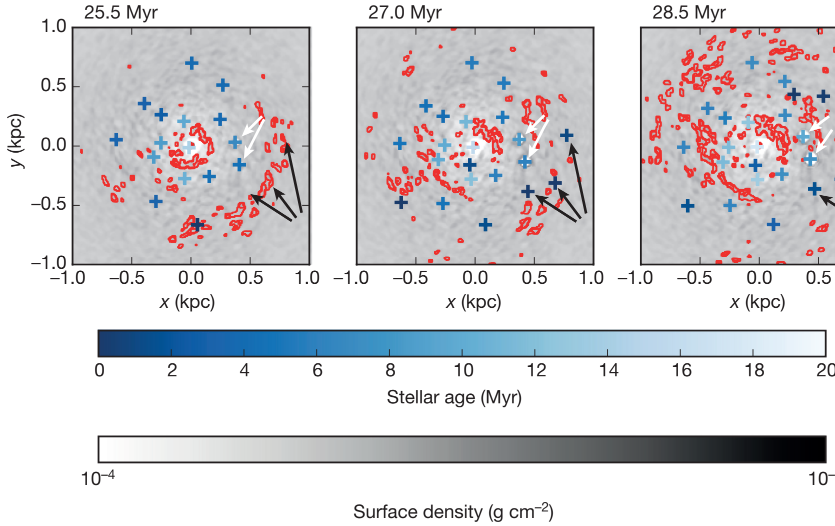


Figure 4 | The effect of photoelectric heating. Snapshots of the 5-kpc SN + PE simulation in 1.5-Myr intervals showing the surface density of gas in grayscale. The red contours show regions where the density is at least 80% of the threshold density for forming stars at the temperature of the gas (the Jeans density). Star particles formed within the 20 Myr preceding each snapshot are shown as crosses, with darker colours indicating younger stars. The stars indicated by black arrows, which formed between the first and second snapshots, have heated nearby gas by the third snapshot, preventing further star formation. Supernova remnants (indicated by white arrows) have clear morphological signatures, but do not substantially affect star formation.

heating and supernovae, but it is the moderate, local, instantaneous, volumetric heating of the former that controls the depletion time. Supernovae, despite their ability to eject mass from the galaxy at a rate larger than the star formation rate, are unable to shut down star formation locally and instantaneously. In the long run, even though dwarf galaxies have 10–100 Gyr depletion times, large mass loading factors mean that the timescale on which gas is lost from the interstellar medium is shorter—between 1 Gyr and 10 Gyr—implying that these galaxies may be in equilibrium between gas inflow and star formation plus outflows[25–27]. Therefore, the star formation rate in these galaxies would ultimately be set by the value of the mass loading factor, whereas the mass of gas in the interstellar medium would be set by the depletion time. In turn, the mass loading factor is set by supernovae and the depletion time by photoelectric heating.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 October 2015; accepted 19 April 2016.
Published online 27 June 2016.

1. Draine, B. T. Photoelectric heating of interstellar gas. *Astrophys. J. Suppl. Ser.* **36,** 595–619 (1978).
2. Bekki, K. Dust-regulated galaxy formation and evolution: a new chemodynamical model with live dust particles. *Mon. Not. R. Astron. Soc.* **449,** 1625–1649 (2015).
3. Tasker, E. J. Star formation in disk galaxies. II. The effect of star formation and photoelectric heating on the formation and evolution of giant molecular clouds. *Astrophys. J.* **730,** 11 (2011).
4. Tasker, E. J., Wadsley, J. & Pudritz, R. Star formation in disk galaxies. III. Does stellar feedback result in cloud death? *Astrophys. J.* **801,** 33 (2015).
5. Hu, C.-Y., Naab, T., Walch, S., Glover, S. C. O. & Clark, P. C. Star formation and molecular hydrogen in dwarf galaxies: a non-equilibrium view. *Mon. Not. R. Astron. Soc.* **458,** 3528–3553 (2016).
6. Hopkins, P. F., Quataert, E. & Murray, N. Self-regulated star formation in galaxies via momentum input from massive stars. *Mon. Not. R. Astron. Soc.* **417,** 950–973 (2011).
7. Hopkins, P. F., Narayanan, D. & Murray, N. The meaning and consequences of star formation criteria in galaxy models with resolved stellar feedback. *Mon. Not. R. Astron. Soc.* **432,** 2647–2653 (2013).
8. Hayward, C. C. & Hopkins, P. F. How stellar feedback simultaneously regulates star formation and drives outflows. Preprint at http://arxiv.org/abs/1510.05650 (2015).
9. Krumholz, M. R., McKee, C. F. & Tumlinson, J. The atomic-to-molecular transition in galaxies. II: H I and H_2 column densities. *Astrophys. J.* **693,** 216–235 (2009).
10. Krumholz, M. R. & Dekel, A. Metallicity-dependent quenching of star formation at high redshift in small galaxies. *Astrophys. J.* **753,** 16 (2012).
11. Christensen, C. *et al.* Implementing molecular hydrogen in hydrodynamic simulations of galaxy formation. *Mon. Not. R. Astron. Soc.* **425,** 3058–3076 (2012).
12. Krumholz, M. R. The star formation law in molecule-poor galaxies. *Mon. Not. R. Astron. Soc.* **436,** 2747–2762 (2013).
13. Makiya, R., Totani, T., Kobayashi, M. A. R., Nagashima, M. & Takeuchi, T. T. Galaxy luminosity function and its cosmological evolution: testing a new feedback model depending on galaxy-scale dust opacity. *Mon. Not. R. Astron. Soc.* **441,** 63–72 (2014).
14. Hunter, D. A. *et al.* Little things. *Astron. J.* **144,** 134 (2012).
15. Bigiel, F. *et al.* A constant molecular gas depletion time in nearby disk galaxies. *Astrophys. J.* **730,** L13 (2011).
16. Bryan, G. L. *et al.* ENZO: an adaptive mesh refinement code for astrophysics. *Astrophys. J. Suppl. Ser.* **211,** 19 (2014).
17. Papastergis, E., Cattaneo, A., Huang, S., Giovanelli, R. & Hanes, M. P. A direct measurement of the baryonic mass function of galaxies and implications for the galactic baryon fraction. *Astrophys. J.* **759,** 138 (2012).
18. Okamoto, T., Gao, L. & Theuns, T. Mass loss of galaxies due to an ultraviolet background. *Mon. Not. R. Astron. Soc.* **390,** 920–928 (2008).
19. Dekel, A. & Silk, J. The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. *Astrophys. J.* **303,** 39–55 (1986).
20. Janowiecki, S. *et al.* (Almost) dark HI Sources in the ALFALFA survey: the intriguing case of HI1232+20. *Astrophys. J.* **801,** 96 (2015).
21. Broeils, A. H. & Rhee, M.-H. Short 21-cm WSRT observations of spiral and irregular galaxies. H I properties. *Astron. Astrophys.* **324,** 877–887 (1997).
22. Katz, N. Dissipational galaxy formation. II. Effects of star formation. *Astrophys. J.* **391,** 502–517 (1992).
23. Turk, M. J. *et al.* yt: a multi-code analysis toolkit for astrophysical simulation data. *Astrophys. J. Suppl. Ser.* **192,** 9 (2011).
24. Cannon, J. M. *et al.* The ALFALFA "Almost Darks" campaign: pilot VLA H I observations of five high mass-to-light ratio systems. *Astron. J.* **149,** 72 (2015).
25. Davé, R., Finlator, K. & Oppenheimer, B. D. An analytic model for the evolution of the stellar, gas and metal content of galaxies. *Mon. Not. R. Astron. Soc.* **421,** 98–107 (2012).
26. Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A. & Peng, Y. Gas regulation of galaxies: the evolution of the cosmic specific star formation rate, the metallicity-mass-star-formation rate relation, and the stellar content of halos. *Astrophys. J.* **772,** 119 (2013).
27. Forbes, J. C., Krumholz, M. R., Burkert, A. & Dekel, A. On the origin of the fundamental metallicity relation and the scatter in galaxy scaling relations. *Mon. Not. R. Astron. Soc.* **443,** 168–185 (2014).

Acknowledgements J.C.F. and M.R.K. acknowledge support from Hubble Archival Research grant HST-AR-13909. This work was also supported by NSF grants AST-09553300 and AST-1405962, NASA ATP grant NNX13AB84G and NASA TCAN grant NNX14AB52G (J.C.F., M.R.K. and N.J.G.), and by Australian Research Council grant DP160100695. A.D. acknowledges support from the grants ISF 124/12, I-CORE Program of the PBC/ISF 1829/12, BSF 2014-273 and NSF AST-1405962. Simulations were carried out on NASA Pleiades and the UCSC supercomputer Hyades, supported by NSF grant AST-1229745.

Author Contributions J.C.F. and N.J.G. developed modifications to the publicly available Enzo code used in this work. The code was run and the results were analysed by J.C.F. The manuscript was written by J.C.F. and edited by all authors. The work was supervised and routinely advised by M.R.K. and A.D.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C.F. (jcforbes@ucsc.edu).

Reviewer Information *Nature* thanks R. Makiya and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Our simulations follow the evolution of an isolated dwarf galaxy for approximately 100 Myr using the adaptive mesh refinement code Enzo[16]. We use a piecewise-parabolic mesh hydrodynamics solver, with an HLLC Riemann solver to follow the motion of the gas. The gravitational potential is computed on the same mesh used to solve the hydrodynamics. Dark matter and stars are included as collisionless particles acting independently of the hydrodynamics, except when new particles are formed or existing particles inject mass, metals and energy back into the gas according to our feedback model.

We use the same sort of initialization as in previous work[28], although with parameters appropriate for a dwarf galaxy. In particular, we create a set of stellar and dark matter particles with the MakeGalaxy code[29]. The density of a stellar disk and bulge are pre-specified, and their velocities are set so that the galaxy begins in approximate Jeans equilibrium. The dark matter has an initial Hernquist profile. We use a halo mass of $10^{10}M_\odot$, a concentration of 10, a spin parameter of 0.04, a stellar mass of $10^7M_\odot$ and a stellar scale-length of 300 pc.

We initialize the density of the gas disk as a function of cylindrical radius r and height z according to

$$\rho_{\rm d}(r,z)=\rho_0\exp\left(-\frac{r}{r_{\rm d}}\right)\exp\left(-\frac{|z|}{H}\right) \qquad (1)$$

where ρ_0 is a normalizing constant chosen to set the initial gas mass of the disk, $r_{\rm d}$ is the scale-length and H is the scale-height. Equation (1) applies until the putative disk pressure $\rho_{\rm d}T_{\rm d}$ (where $T_{\rm d}$ is the temperature of the disk) falls below the halo pressure $\rho_{\rm h}T_{\rm h}$, at which point the density and temperature of each cell are set to those of the halo, $\rho_{\rm h}$ and $T_{\rm h}$, respectively. Although, $\rho_{\rm d}$ is spatially dependent, $\rho_{\rm h}$, $T_{\rm d}$ and $T_{\rm h}$ are all taken to be constant. In these simulations we set $T_{\rm h}=10^6$ K, $T_{\rm d}=1{,}000$ K, $\rho_{\rm h}=2.34\times10^{-30}$ g cm^{-3} and $\rho_0=1.41\times10^{-24}$ g cm^{-3}. The scale-length and scale-height of the disk are $r_{\rm d}=5$ kpc for most simulations and $H=30$ pc, respectively. A number of simulations in the main text use $r_{\rm d}=1$ kpc. These initial conditions are chosen to minimize the time the disk spends in its initial collapse phase and the influence of the galactic halo on the dynamics. In particular, at this temperature the halo does not monolithically cool onto the galaxy over the timescale on which the simulation is run.

These initial conditions are evolved under the influence of hydrodynamics, gravity and cooling, with additional subgrid prescriptions for the creation of new star particles and feedback from young stars, specifically stellar winds, type II supernovae and a rough treatment of heating from photoionization. We also include a prescription for self-consistently calculating the volumetric heating rate from the grain photoelectric effect.

Feedback. Our feedback prescription is designed to avoid the *ad hoc* modifications to the physics that are typically necessary to produce feedback strong enough to affect the properties of the galaxy. Our goal is to show that none of these modifications is necessary to have effective feedback at sufficiently high resolution. The basic physical picture is that once the adiabatic radius of the supernovae can be resolved, they will expand in a somewhat realistic way, rather than immediately losing all their energy to the cell, as in the famous overcooling problem[22].

When each star particle is formed in the simulation with initial mass $M_{\rm p}$, we draw a number from a Poisson distribution to determine how many type II supernovae will be produced by the particle. The rate parameter of this Poisson distribution is taken to be $\lambda=1{,}104M_{\rm p}/(10^6M_\odot)$, where the prefactor is the number of supernovae produced over the lifetime of a $10^6M_\odot$ stellar population formed in a burst of star formation according to Starburst99 (ref. 30) using a Chabrier initial mass function[31].

For each supernova produced, we next draw a number from the delay-time distribution of type II supernovae, again determined from the output of Starburst99. This is essentially just the convolution of the lifetime of massive stars with the initial mass function. At the hydrodynamical time step during which the supernova explodes, we add 10^{51} erg to the internal energy of the cell in which the supernova resides. The mass of ejecta and its metallicity are also determined from the Starburst99 output, and are fit to piecewise-polynomial functions for use in Enzo. To ensure that the supernova goes off on the highest level of refinement in the simulation, we make the star particles 'must refine' if the particle has any supernovae remaining in its future. This means that the cells containing such particles are marked for refinement to the highest resolution level, guaranteeing that they, and at least the surrounding two cells in each direction, will be on the highest resolution level.

In addition to the energy from the supernovae added to the cell at the appropriate time step, we also add energy to the cell before the supernova itself. From the delay time for each supernovae, we infer the mass of the star that will be exploding, from which we can estimate the total ionizing luminosity for the particle by adding up the contribution from every massive star that has yet to explode in that particle. Given an ambient density, that is, the density in the surrounding gas cell, we may then estimate the volume of gas that can be ionized by the calculated luminosity, as in the standard Strömgren analysis. If this volume is larger than the cell, and the cell has a temperature below 10^4 K, internal energy is added to the cell to set the temperature to 10^4 K. The cell is not modified by this photoionization prescription if its temperature is above 10^4 K.

For cells with Strömgren volumes smaller than the cell volume, we compute the internal energy of the Strömgren sphere and compare it to the internal energy across the entire volume of the cell. If the whole cell has less internal energy than the Strömgren sphere alone, then we add internal energy to the cell to make the two equal; otherwise we do nothing.

This treatment of photoionization feedback is extremely conservative. In the limit of small cell sizes or low densities, we may vastly underestimate the size of the H II region, because only one cell will be at 10^4 K. In the limit of high density or large cells, the total energy being injected into the gas is also conservative, in that the energy of the cell is changed only if the entire internal energy of the cell is smaller than the internal energy associated with the Strömgren sphere on its own.

In addition to photoionization, we include feedback from stellar winds. Once again we rely on Starburst99 to compute the wind luminosity, with several small modifications. The specific energy of the wind is taken from Starburst99 only when the particle will in the future experience a supernova from the death of a star with a delay time less than 10 Myr. If no such supernovae will explode in the future, then the cell still loses mass to winds, but their specific energy is set to $(10\,{\rm km\,s^{-1}})^2$. This bifurcation in wind temperatures takes place in the Universe as a result of a sharp change in opacity in the atmospheres of stars as a function of mass.

One additional change we make to the specific energy of the wind occurs right at the beginning of the life of the star particle. During this time, the most massive stars in the population drive extremely hot winds, exceeding 10^9 K. When the outflowing winds dominate the material in a single cell, as sometimes happens in our simulations, this high temperature can substantially slow down the simulation and cause other numerical problems associated with a high density contrast. To ameliorate this issue we cap the wind specific energy at $10^{8.5}$ K. We keep the total wind energy injected constant by slightly increasing the mass lost during this time. This changes the total mass returned by only a tiny amount because this phase is so short.

Extended Data Fig. 1 shows that every supernova in the 10-pc-resolution simulation with the 5-kpc initial scale-length explodes in a cell with a density below the critical density at which a supernova remnant would cool before expanding to be the size of a single cell in the simulation. We calculate this critical density by adopting the following value for the radius at which a supernova remnant exits the Sedov phase and enters the pressure-driven snowplough (PDS) phase[32]:

$$R_{\rm PDS}=14\frac{E_{51}^{2/7}}{n_{\rm H}^{3/7}(Z/Z_\odot)^{1/7}}\,{\rm pc}$$

We take the energy of the supernova in units of 10^{51} erg to be $E_{51}=1$, and set the mass fraction of elements heavier than helium to a value appropriate for the Sun, $Z=Z_\odot=0.02$. In our initial conditions, the disk component has $Z/Z_\odot=0.1$ and the halo has $Z/Z_\odot=0.01$, but for the purposes of calculating the critical density we use a higher value because supernovae produce enough metals that locally Z may be substantially higher than its initial value. By setting $R_{\rm PDS}$ to the size of a single cell in the simulation, we can solve for the value of $n_{\rm H}$ (the number density of hydrogen atoms in units of cm^{-3}) at which the PDS phase would be marginally resolved.

By resolving this crucial piece of physics in these galaxies, we find that our results are relatively insensitive to the resolution at which we run the simulations. We compare the depletion time for each physical scenario, run at 10-pc, 5-pc and 2.5-pc resolution, in Extended Data Fig. 3, and we find that the results of the simulations tend to become independent of resolution after roughly 100 Myr of evolution. This is less clear for the runs that include supernovae, for which the 2.5-pc simulations have not advanced as far as their lower-resolution counterparts; but even here there is reasonable agreement between the 10-pc and 5-pc runs.

Photoelectric heating. Far-ultraviolet (FUV) photons from young stars liberate electrons from dust grains in the interstellar medium. This is the primary means by which the neutral atomic gas in the interstellar medium is heated in the Milky Way. To include it in our simulation, we assume the following proportionality:

$$\Phi\propto F_{\rm FUV}Zn_{\rm H}\ {\rm erg\,s^{-1}\,cm^{-3}}$$

The heating rate Φ from FUV photons is proportional to their flux $F_{\rm FUV}$, and the density of metals $Zn_{\rm H}$. At low densities and high temperatures, there is an additional dependence on the electron density and gas temperature, but these effects are negligible in the cold, dense gas where FUV heating is important for suppression of star formation, so we omit that effect. We also do not include cosmic ray heating, because this is roughly an order of magnitude less important than FUV heating under the optically thin conditions that prevail in the low density, dust poor galaxies we are simulating.

We calculate the FUV flux in the simulation by taking the luminosity L for each star particle to be

$$\log_{10}L(t_7)=\log_{10}[M_\mathrm{p}/(10^6M_\odot)]+\begin{cases}\sum_{j=0}^{2}p_jt_7^j & \text{if } t_7\le 3\\ \sum_{j=0}^{6}q_jt_7^j & \text{if } 3\le t_7\le 10\\ 0 & \text{otherwise}\end{cases}\qquad(2)$$

where t_7 is the age of the star particle in units of 10^7 yr, and M_p is the mass of the star particle. L is in units of erg s^{-1}. The coefficients of polynomials (p_j and q_j) are given in Extended Data Table 1. This expression is the result of integrating the output spectrum of a Starburst99 single-burst model for a cluster mass of $10^6M_\odot$ (hence the pre-factor in equation (2)) over the range 8–13.6 eV at finely spaced time intervals out to 100 Myr. This function is somewhat sensitive to the initial mass function—recent indications of a bottom-light initial mass function[33] in dwarfs would increase it by a factor of approximately two.

The FUV flux is then simply $F_\mathrm{FUV}=L(t)/(4\pi r^2)$, where r is the three-dimensional distance from the centre of the gas cell in question to the star in question. This expression for F_FUV neglects any effects from self-shielding, which should be negligible in the galaxies we have simulated, given their low column densities and low metallicities. The total FUV flux at a given cell is the sum of this quantity over all stars in the simulation. If a given star's contribution to the FUV flux varies by less than 10% across a given grid (the computational element one step above cells in Enzo), then we approximate the contribution of that star as constant across the grid, to avoid doing the full order N by M computation (where N is the number of cells and M is the number of particles) in regions far from FUV-emitting particles. To compute Φ/n_H, we scale Φ/n_H, F_FUV and Z to the known values of these quantities in the solar neighbourhood:

$$\frac{\Phi}{n_\mathrm{H}}=8.5\times10^{-26}\frac{\mathrm{erg}}{\mathrm{s}}\frac{F_\mathrm{FUV}}{0.0015859021\ \mathrm{erg\ s^{-1}\ cm^{-2}}}\frac{Z}{Z_\odot}$$

The numerical constant in the denominator is the Habing[34] estimate of the intensity of the interstellar radiation field in the solar neighbourhood, multiplied by c (the speed of light) to convert to a flux. The pre-factor of 8.5×10^{-26} is the photoelectric heating rate normalized to the Habing value for the solar neighbourhood from a radially dependent model of the interstellar medium of the Milky Way[35]. This quantity is computed for each cell in the simulation, and fed to the Grackle library, which computes the rate of change of the internal energy density[16,36,37]. An example of Φ/n_H in the 'PE only' simulation is shown in Fig. 3.

This method is a substantial improvement over many current implementations of the interstellar radiation field. Many simulations do not include this source of diffuse heating at all. Some include it as constant throughout the simulation volume[3,4], sometimes with a correction for self-shielding[5], but without regard to the individual sources or time-dependence of the FUV photons. Other simulations explicitly approximate the radiative transfer of these photons[38], while others go even further and explicitly model the dust particles via which this radiation interacts with the gas[2,39,40]. Although the last two methods are better approximations to the interstellar radiation field than what we have implemented here, our simulations have a higher resolution. As a result, we resolve the Sedov–Taylor phase of the supernovae (Extended Data Fig. 1) and, hence, can definitively show whether supernova feedback or the interstellar radiation field is dominant in the regulation of star formation in dwarf galaxies.

Star formation. The star formation prescription we use is similar to many commonly used schemes. At each time step, the probability of forming a star in each cell is taken to be

$$p_*=\begin{cases}0 & \text{if excluded}\\ \min\left(1,\varepsilon_\mathrm{ff}\dfrac{\mathrm{d}t}{t_\mathrm{ff}}\dfrac{M_\mathrm{cell}}{M_\mathrm{p}}\right) & \text{otherwise}\end{cases}$$

where $\mathrm{d}t$ is the simulation time step (at the refinement level in question), ε_ff is the efficiency of star formation per free fall time, M_cell is the mass of gas in the computational cell, and the free fall time $t_\mathrm{ff}=\sqrt{3\pi/(32G\rho)}$. Cells are excluded if their mass is less than the Jeans mass in that cell or if the cell is not on the maximum refinement level. In other words, cells must have a density

$$\rho>\frac{\gamma\pi k_\mathrm{B}T}{N_\mathrm{J}^2G\mu m_\mathrm{H}(\Delta x)^2}$$

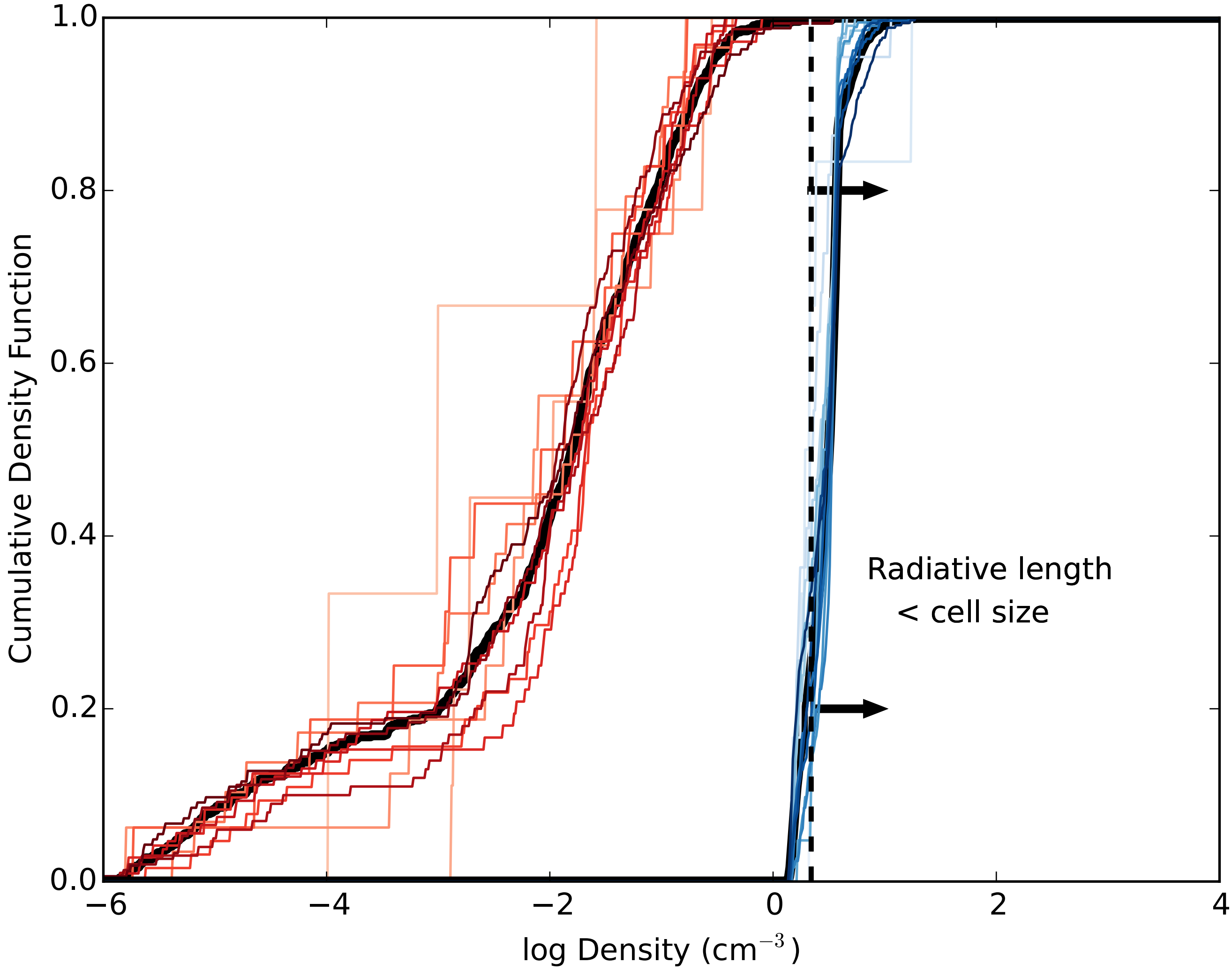
to form stars. Here $\gamma=5/3$ is the ratio of specific heats, k_B is the Boltzmann constant, μ is the mean molecular weight in units of the hydrogen mass m_H, Δx is the size of the cells on the maximum refinement level, G is the gravitational constant and $N_\mathrm{J}=4$ is the number of cells by which we require that the Jeans length be resolved throughout the simulation to avoid artificial fragmentation[41]. This criterion corresponds to a diagonal line in the density–temperature phase diagram above which stars are not allowed to form (Extended Data Fig. 4). The value of p_* is such that the average star formation rate in the non-excluded cells follows a volumetric Schmidt Law: $\dot\rho=\varepsilon_\mathrm{ff}\rho/t_\mathrm{ff}$. This model has a few parameters that we must set, although we argue that we do not have a huge amount of freedom to change them.

The efficiency per free fall time is constrained by observations to be within a factor of a few of 0.01 (refs 42, 43), so we simply adopt this central value. The mass of individual star particles M_p should in principle be low enough to be irrelevant—this would have the advantage of sampling the star formation rate density very well and forming stars if and only if the cell were Jeans unstable. This introduces two numerical difficulties. The first is that the Jeans mass for the coldest gas in our simulations can be comparable to the mass of a single star. It would be problematic for our feedback recipe to form such low-mass stars, in that a $120M_\odot$ star or a supernova that ejects $10M_\odot$ of material could never exist within a $10M_\odot$ particle. Moreover, the number of star particles we would have to follow increases as M_p^{-1}. Each of these issues would be avoided by choosing a large particle mass. However, using a large mass increases the chance that the cell in which the star formed would have insufficient mass to supply all of the gas needed to form the particle. When this happens, to maintain global mass conservation, we gather the mass from neighbouring cells. Each cell supplying gas contributes the same fraction of its mass to the new particle. We have found that a particle mass of $M_\mathrm{p}=50M_\odot$ constitutes a reasonable balance between keeping star formation local to one cell and keeping the supernova ejecta mass less than the mass of a single star particle.

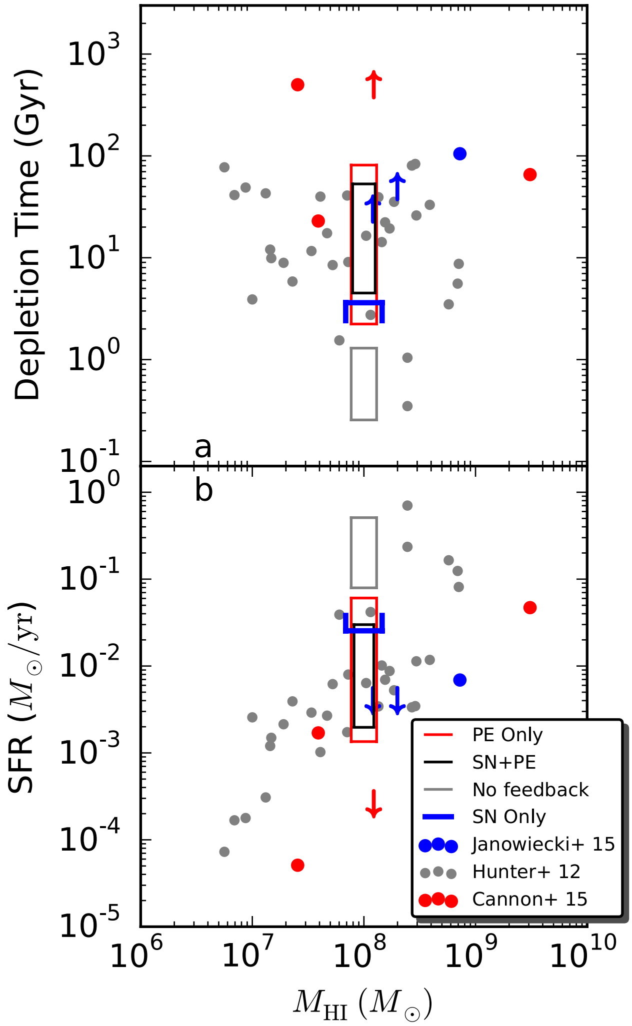
Comparison to observations. In Extended Data Fig. 2, we compare the star formation rate and depletion time of the simulations to those of dwarf galaxies from three different samples[14,20,24]. We exclude four blue, compact dwarf galaxies from this sample, because these are known to be undergoing starbursts. We do not include estimates for the errors on each point. The statistical standard errors are of the order of 10%, but this underpredicts the true error, which is dominated by systematic uncertainty in distance (H I mass and star formation rate) and the assumptions made in converting ultraviolet or Hα luminosities to star formation rates[44]. For the two smaller galaxies in ref. 20, no strict upper limits are given on the star formation rate; but, because they are not detected in the same Hα image in which a star formation rate of $4\times10^{-3}M_\odot$ yr^{-1} was measured for the largest galaxy in the sample, we take this as a very conservative upper limit on the star formation rate for these two galaxies (S. Janowiecki, private communication).

Data availability. The numerical experiments presented here were run with a fork of the Enzo code, available from https://bitbucket.org/jforbes/enzo-dev-jforbes, in particular, change set *daed04d1e5e6*. This altered version of Enzo also requires an altered version of the Grackle cooling library, available from https://bitbucket.org/jforbes/grackle, in particular, change set *12d3856*. A subset of the raw data files is available from http://www.johncforbes.com/dwarfs.html.

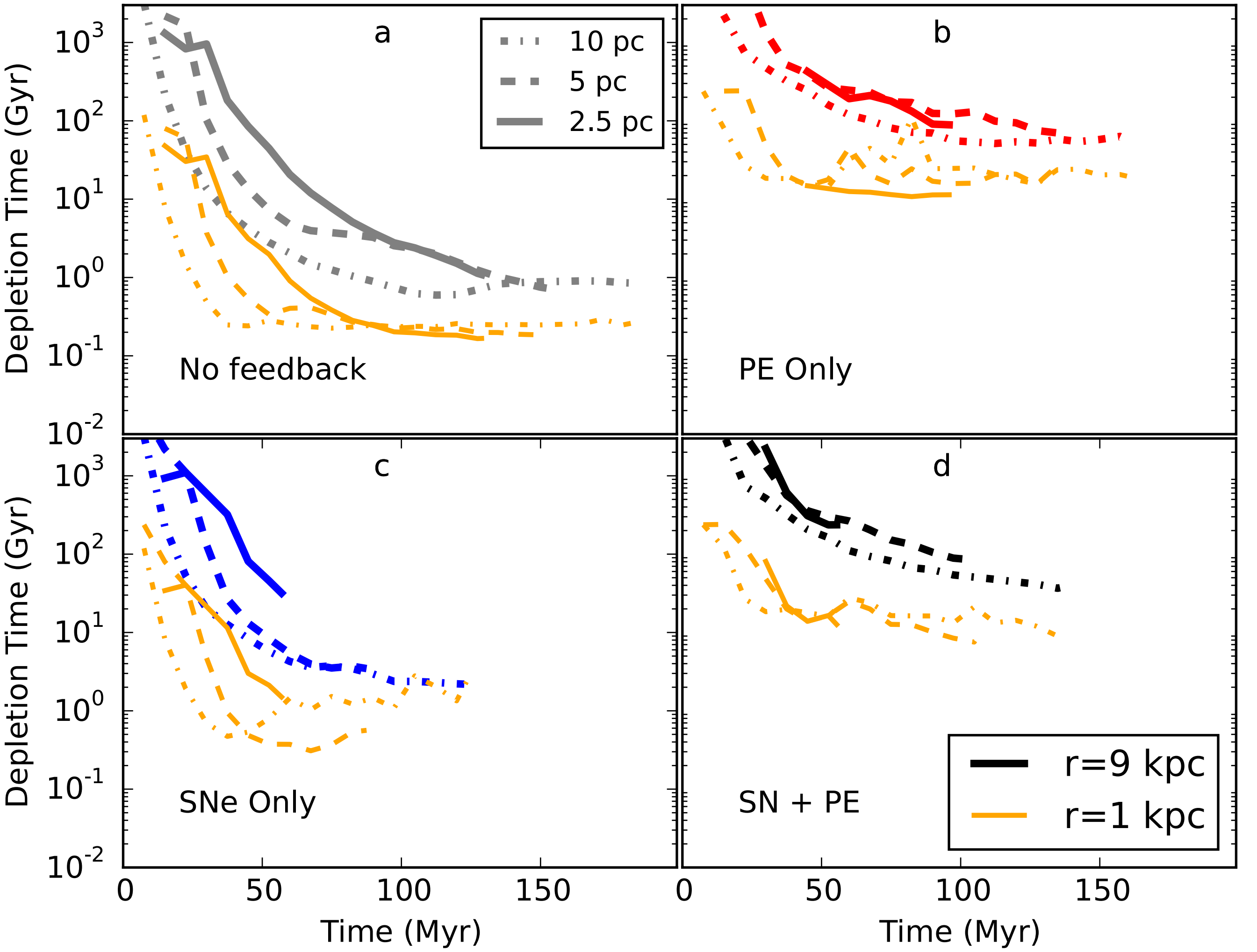
28. Goldbaum, N. J., Krumholz, M. R. & Forbes, J. C. Mass transport and turbulence in gravitationally unstable disk galaxies. I. The case of pure self-gravity. *Astrophys. J.* **814,** 131 (2015).
29. Springel, V. & White, S. D. M. Tidal tails in cold dark matter cosmologies. *Mon. Not. R. Astron. Soc.* **307,** 162–178 (1999).
30. Leitherer, C. *et al.* Starburst99: synthesis models for galaxies with active star formation. *Astrophys. J. Suppl. Ser.* **123,** 3–40 (1999).
31. Chabrier, G. Galactic stellar and substellar initial mass function. *Publ. Astron. Soc. Pac.* **115,** 763–795 (2003).
32. Thornton, K., Gaudlitz, M., Janka, H.-T. & Steinmetz, M. Energy input and mass redistribution by supernovae in the interstellar medium. *Astrophys. J.* **500,** 95–119 (1998).
33. Geha, M. *et al.* The stellar initial mass function of ultra-faint dwarf galaxies: evidence for IMF variations with galactic environment. *Astrophys. J.* **771,** 29 (2013).
34. Habing, H. J. The interstellar radiation density between 912 Å and 2400 Å. *Bull. Astron. Inst. Netherlands* **19,** 421–431 (1968).
35. Wolfire, M. G., McKee, C. F., Hollenbach, D. & Tielens, A. G. G. M. Neutral atomic phases of the interstellar medium in the galaxy. *Astrophys. J.* **587,** 278–311 (2003).
36. Kim, J.-h. *et al.* The AGORA high-resolution galaxy simulations comparison project. *Astrophys. J. Suppl. Ser.* **210,** 14 (2014).
37. The Grackle Library. https://grackle.readthedocs.org/.
38. Oñorbe, J. *et al.* Forged in FIRE: cusps, cores and baryons in low-mass dwarf galaxies. *Mon. Not. R. Astron. Soc.* **454,** 2092–2106 (2015).
39. Bekki, K. Formation of ultra-compact blue dwarf galaxies and their evolution into nucleated dwarfs. *Astrophys. J.* **812,** L14 (2015).
40. Bekki, K. Formation of emission line dots and extremely metal-deficient dwarfs from almost dark galaxies. *Mon. Not. R. Astron. Soc.* **454,** L41–L45 (2015).
41. Truelove, J. K. *et al.* Self-gravitational hydrodynamics with three-dimensional adaptive mesh refinement: methodology and applications to molecular cloud collapse and fragmentation. *Astrophys. J.* **495,** 821–852 (1998).
42. Krumholz, M. R. & Tan, J. C. Slow star formation in dense gas: evidence and implications. *Astrophys. J.* **654,** 304–315 (2007).
43. Krumholz, M. R., Dekel, A. & McKee, C. F. A universal, local star formation law in galactic clouds, nearby galaxies, high-redshift disks, and starbursts. *Astrophys. J.* **745,** 69 (2012).
44. Chomiuk, L. & Povich, M. S. Toward a unification of star formation rate determinations in the Milky Way and other galaxies. *Astron. J.* **142,** 197 (2011).



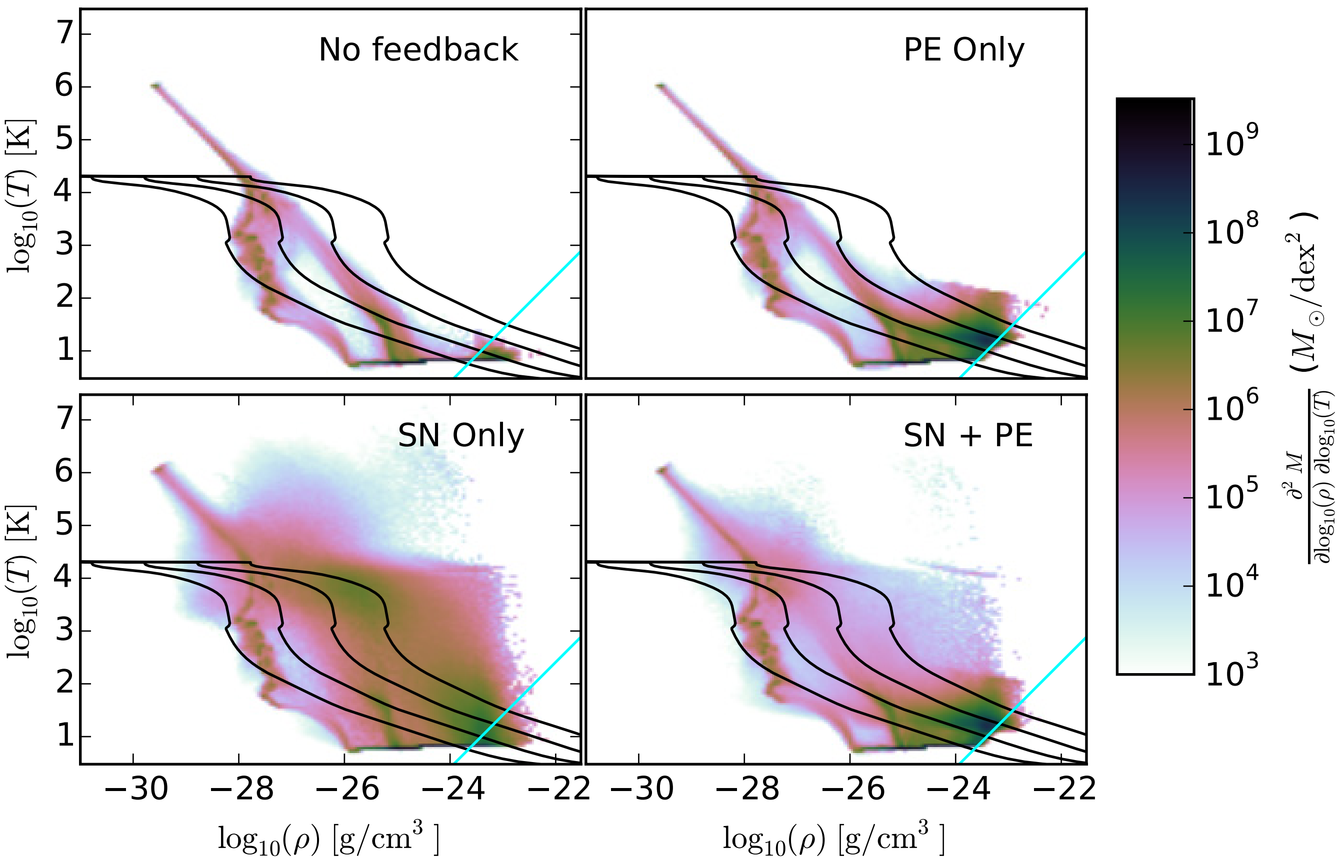
Extended Data Figure 1 | Supernovae are well-resolved. The cumulative distribution of the density of cells in which supernovae explode is shown in red, and that in which stars form is shown in blue. The thin lines show these distributions in different 10-Myr intervals, while the thick lines show the cumulative density function over the entire simulation. The vertical dashed line indicates the density at which a supernova remnant would radiate all of its energy before it expanded to the size of a single cell (10 pc) in the simulation, assuming solar metallicity. Nearly every supernova in the simulation occurs to the left of this line, indicating that the simulation does not suffer from the overcooling problem.



Extended Data Figure 2 | Comparison to observations. a, b, Star formation properties (**a**, depletion time; **b**, star formation rate (SFR)) for a heterogeneous set[14,20,24] of galaxies are shown as a function of gas mass $M_{\rm HI}$. Upper and lower limits are indicated with arrows. Boxes representing the range of values covered by our simulations are over-plotted. The vertical range of the box is determined by the final snapshot for each simulation; the high (low) SFR extremum represents the 1-kpc (5-kpc) disk. No 'SN only' simulation was run for the 1-kpc case, so the blue box is not closed. Only simulations that include photoelectric heating agree with the depletion times observed for bulk of galaxies in the mass range we simulated. See Methods section 'Comparison to observations' for more details.



Extended Data Figure 3 | A resolution study. a–d, The depletion time of all 12 simulations with 5-kpc gas scale-length is plotted over time. The four feedback models are shown (**a**, 'no feedback', grey; **b**, 'PE only', red; **c**, 'SN only', blue; **d**, 'SN + PE', black), with the line style indicating the result for different resolutions (see legend). The orange lines show the depletion time when the measurement is carried out in a cylinder with 1-kpc radius; the other lines use a 9-kpc radius. Regardless of the aperture, the simulations quickly converge; differences between simulations with factor-of-two differences in resolution are small compared to the differences resulting from changing the physics.



Extended Data Figure 4 | Phase diagrams after 90 Myr of evolution. Runs with different feedback models and 5-kpc gas scale-length are shown, all at 10-pc resolution. The light blue diagonal lines show the threshold for star formation, at which the gas becomes Jeans unstable on the highest refinement level. The black lines trace where the net cooling rate is zero, assuming different values for the volumetric heating rate, from 10^{-26} erg s^{-1} (highest line) to 10^{-29} erg s^{-1} (lowest line). Photoelectric heating raises the typical temperature of gas near the star formation threshold such that moderate star formation can stabilize nearby gas against collapse.

Extended Data Table 1 | Parameters for a fit

j	p_j	q_j
0	41.5709926107	40.7875024388
1	2.40501751872	-0.227682606645
2	-9.19544984847	0.0078916423535
3	10.5203892767	0
4	-5.72637964222	0
5	1.50479521662	0
6	-0.153355377095	0

Coefficients for the piecewise-polynomial fit to the FUV luminosity of a simple stellar population as a function of its age (Methods, equation (2)).

doi:10.1038/nature18638

Solar-type dynamo behaviour in fully convective stars without a tachocline

Nicholas J. Wright[1] & Jeremy J. Drake[2]

In solar-type stars (with radiative cores and convective envelopes like our Sun), the magnetic field powers star spots, flares and other solar phenomena, as well as chromospheric and coronal emission at ultraviolet to X-ray wavelengths. The dynamo responsible for generating the field depends on the shearing of internal magnetic fields by differential rotation[1,2]. The shearing has long been thought to take place in a boundary layer known as the tachocline between the radiative core and the convective envelope[3]. Fully convective stars do not have a tachocline and their dynamo mechanism is expected to be very different[4], although its exact form and physical dependencies are not known. Here we report observations of four fully convective stars whose X-ray emission correlates with their rotation periods in the same way as in solar-type stars. As the X-ray activity–rotation relationship is a well-established proxy for the behaviour of the magnetic dynamo, these results imply that fully convective stars also operate a solar-type dynamo. The lack of a tachocline in fully convective stars therefore suggests that this is not a critical ingredient in the solar dynamo and supports models in which the dynamo originates throughout the convection zone.

Stars across the Hertzsprung–Russell diagram are known to emit X-rays, with only a few exceptions. In main-sequence solar-type and low-mass stars, the X-rays arise from a magnetically confined plasma known as a corona that reaches temperatures of several million kelvin[5]. Coronal X-ray emission is ultimately powered by the dissipation of the magnetic fields generated by an interior magnetic dynamo. A close relation between the surface magnetic flux and X-ray radiance based on both solar and stellar observations that span several decades for both quantities[6] indicates that X-ray emission is a reliable proxy for magnetic activity. The dynamo is thought to be driven in part by differential rotation in the interior of the star[1,2], which itself is generated by the action of the Coriolis force on the rotating convective envelope, but the detailed mechanism remains to be properly understood[7]. The relationship between stellar rotation and tracers of magnetic activity is therefore an important probe of the stellar dynamo.

In solar-type stars, X-ray emission is observed to increase monotonically with increasing stellar rotational velocity for periods exceeding a few days[8,9]. This relationship is often quantified in terms of the dependence of the ratio of stellar luminosity expended in X-rays to bolometric luminosity $L_X/L_{\rm bol}$, on the Rossby number, which is defined as $\mathrm{Ro} = P_{\rm rot}/\tau$, the ratio of the stellar rotation period and the mass-dependent convective turnover time[10]. A recent study[9] using the largest available sample of 824 solar- and late-type stars fitted the rotation–activity relation as $L_X/L_{\rm bol} = 5.3 \times 10^{-6}\mathrm{Ro}^{-2.7}$.

This relationship has been observed in stars from late F-type through to early M-type, that is, those with radiative cores and convective envelopes. The interface layer between these two regions, named the tachocline, is believed to play an important role in the generation of the magnetic field[3]. Shear between the rigidly rotating core and the differential rotation of the convective envelope with latitude is thought to amplify and store the magnetic field[11], generating what is known as an α–Ω dynamo, named for the interplay between cyclonic eddies (the α effect) and the shearing of the field (the Ω effect).

Others have argued that the latitudinal and radial gradients of the angular velocity in the convection zone may be sufficient for global dynamo action[12,13]. In the Sun, the extremely strong levels of radial shear just beneath the surface are actually higher than in the tachocline, making it plausible that the solar dynamo is distributed across the convection zone rather than confined to the tachocline[12].

Despite this hypothesis it has been widely accepted that magnetic structures in the convection zone would be disrupted by magnetic buoyancy or turbulent pumping, preventing large-scale magnetic fields from being established there[14]. However, some recent three-dimensional magneto-hydrodynamic simulations without a tachocline have produced persistent magnetic wreaths in the convection zone[15] and shown that it is possible to produce large-scale magnetic fields in stellar convection layers[16,17].

For very fast rotators, the rotation–activity relationship has been found to break down, with X-ray luminosity reaching a saturation level of approximately $L_X/L_{\rm bol} \approx 10^{-3}$, independent of the spectral type[18]. This saturation level is reached at a Rossby number of approximately 0.13 ± 0.02 (ref. 9; the error represents one standard deviation), corresponding to a rotation period that increases towards later spectral types, from 2 days for a star similar to the Sun to up to about 20 days for low-mass M dwarfs, and is also seen in both chromospheric emission and magnetic field measurements. It is unclear whether this is caused by a saturation of either the dynamo mechanism or the transport of the magnetic flux to the corona, a change in the type of dynamo at work within the star[9] or because coronal X-ray emission itself becomes insensitive to the strength of the magnetic field, the energy of which is then dissipated in other ways.

Main-sequence stars later than spectral type M3–3.5 ($M < 0.4M_\odot$, where M is the mass of the star and $M_\odot$ is the mass of the Sun) are predicted to be fully convective and therefore do not possess a tachocline. If the tachocline is critical to the operation of a solar-type dynamo, fully convective stars should not be able to sustain such a dynamo. Instead, it is generally thought that they generate magnetic fields entirely by helical turbulence[4]. Nevertheless, observations indicate that stars throughout the M-type spectral range exhibit high magnetic field strengths[19] and high fractional X-ray luminosities[9]. In fact, no discernible difference in magnetic activity properties has been identified on either side of the fully convective boundary.

One problem with existing studies is that nearly all fully convective stars that have so far been studied have saturated levels of X-ray emission[9]. It is therefore unclear whether these stars all exhibit saturated levels of X-ray emission (possibly hinting at some facet of their dynamo mechanism), or whether slower rotators follow a more solar-like rotation–activity relationship. Studies of magnetic activity in slowly rotating fully convective stars have so far been lacking, adding to the uncertainty about their dynamo mechanism. Despite this, fully convective stars are common and, given their spin-down times of a few billion years[20], at least half of all such stars are expected to be slow rotators.

[1]Astrophysics Group, Keele University, Keele ST5 5BG, UK. [2]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

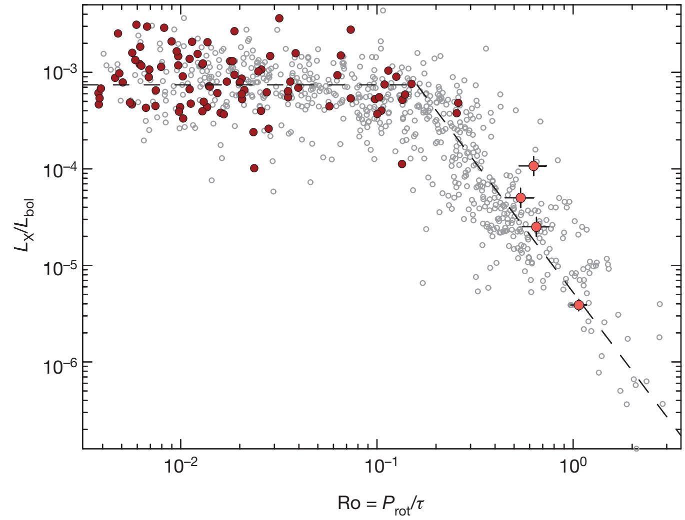


Figure 1 | Rotation–activity relationship diagram for partly and fully convective stars. Fractional X-ray luminosity, L_X/L_{bol}, plotted against the Rossby number, $Ro = P_{rot}/\tau$, for 824 partly (grey circles) and fully (dark red circles) convective stars from the most recent large compilation of stars with measured rotation periods and X-ray luminosities[7]. The best-fitting saturated (horizontal) and unsaturated (diagonal) rotation–activity relationships from that study are shown as black dashed lines. The four slowly rotating fully convective M dwarfs studied here are shown in light red (error bars indicate one standard deviation). The uncertainties for the other data points are not quantified, but will be comparable to the M dwarfs for the Rossby number and approximately twice as large for L_X/L_{bol}.

Understanding the dynamo mechanism in these slowly rotating stars is important for various astrophysical problems, including the dynamo-driven angular momentum loss rate of low-mass stars, the particle and photon radiation environment of exoplanets and the notorious period gap in the cataclysmic variables. The lack of cataclysmic variables with periods in the 2–3 h range is often attributed to a change in rotational spin-down for fully convective stars with diminished magnetic dynamos and stellar winds[21]. Recent numerical simulations of stellar winds suggest that increasing the complexity of the surface magnetic morphology can also suppress angular momentum loss and spin-down without requiring any change in the total surface magnetic flux[22].

Four slowly rotating, fully convective M-type stars were observed by either NASA's Chandra X-ray Observatory or the ROSAT satellite. These stars are of spectral type M4–5.5, well beyond the fully convective boundary (M3–3.5) and are therefore genuine fully convective stars. Figure 1 illustrates the traditional rotation–activity diagram showing the fractional X-ray luminosity as a function of the Rossby number for all of the stars from the most recent large-scale study of the rotation–activity relationship in solar-type and low-mass stars[9]. The positions of the four slowly rotating fully convective stars studied in this work are also shown, and are in excellent agreement with the rotation–activity relationship of partly convective stars.

The observations presented here provide clear evidence for unsaturated X-ray emission in fully convective stars and quantification of their rotation–activity relationship. The results show that fully convective stars, at least when they have spun down sufficiently, operate a dynamo that exhibits a rotation–activity relationship that is indistinguishable from that of solar-type stars. As the dynamo action in fully convective stars is expected to be different from that in solar-type stars owing to their lack of a tachocline[4], where the magnetic field is thought to be amplified by radial shear, this is a surprising finding.

The most direct conclusion from these observations is that both partly and fully convective stars operate very similar rotation-dependent dynamos in which the tachocline is not a vital ingredient and differential rotation combined with the action of the Coriolis force is sufficient[7]. This implies that current models for the solar dynamo that rely on the tachocline layer to amplify the magnetic field are incorrect, and lends weight to recent three-dimensional magneto-hydrodynamic simulations without a tachocline that produce large-scale magnetic fields entirely within the convective layers[23,24]. Recent studies with mean-field dynamo models have also suggested that differential rotation in the convection zone may play a greater role in the generation of the magnetic field than does the tachocline[25,26].

Another alternative possibility is that fully convective stars are able to generate a purely turbulent dynamo that exhibits a rotation–activity relationship that is similar to that in partly convective stars, but by a different mechanism that does not rely on a shear layer. Existing dynamo simulations for fully convective stars succeed in generating magnetic fields, but are unable to predict their behaviour as a function of the rotation rate[17]. However, it seems unlikely that both partly and fully convective stars would have the same rotation–activity relationship (requiring both their dynamo efficiency and rotational dependence to behave in the same way) without their dynamo mechanisms sharing a major feature.

A third possibility is that convection in the cores of fully convective stars could be magnetically suppressed[27], leading to the existence of a solar-like tachocline, although some studies suggest that convection would not be completely halted, only made less efficient[28]. Furthermore, the field strengths that are necessary for such a transition are 10^7–10^8 G (refs 28, 29), orders of magnitude larger than the fields thought to exist in the solar interior and at levels that simulations suggest are impossible to maintain[30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 April; accepted 2 June 2016.

1. Parker, E. N. Hydrodynamic dynamo models. *Astrophys. J.* **122,** 293–314 (1955).
2. Babcock, H. W. The Sun's magnetic field. *Annu. Rev. Astron. Astrophys.* **1,** 41–58 (1963).
3. Speigel, E. A. & Zahn, J.-P. The solar tachocline. *Astron. Astrophys.* **265,** 106–114 (1992).
4. Durney, B. R., De Young, D. S. & Roxburgh, I. W. On the generation of the large-scale and turbulent magnetic fields in solar-type stars. *Sol. Phys.* **145,** 207–225 (1993).
5. Vaiana, G. S. *et al.* Results from an extensive Einstein stellar survey. *Astrophys. J.* **244,** 163–182 (1981).
6. Pevtsov, A. A. *et al.* The relationship between X-ray radiance and magnetic flux. *Astrophys. J.* **598,** 1387–1391 (2003).
7. Spruit, H. C. *Theories of the Solar Cycle: a Critical View* 39–54 (Springer, 2011).
8. Pallavicini, R. *et al.* Relations among stellar X-ray emission observed from Einstein, stellar rotation and bolometric luminosity. *Astrophys. J.* **248,** 279–290 (1981).
9. Wright, N. J., Drake, J. J., Mamajek, E. E. & Henry, G. W. The stellar-activity-rotation relationship and the evolution of stellar dynamos. *Astrophys. J.* **743,** 48 (2011).
10. Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K. & Vaughan, A. H. Rotation, convection, and magnetic activity in lower main-sequence stars. *Astrophys. J.* **279,** 763–777 (1984).
11. Charbonneau, P. Solar dynamo theory. *Annu. Rev. Astron. Astrophys.* **52,** 251–290 (2014).
12. Brandenburg, A. Location of the solar dynamo and near-surface shear. In *ASP Conf. Series* Vol. 354 (eds Leibacher, J., Stein, R. F. & Uitenbroek, H.) 121–126 (Astronomical Society of the Pacific, 2006).
13. Guerrero, G. & de Gouveia Dal Pino, E. M. How does the shape and thickness of the tachocline affect the distribution of the toroidal magnetic fields in the solar dynamo? *Astron. Astrophys.* **464,** 341–349 (2007).
14. Parker, E. N. The generation of magnetic fields in astrophysical bodies. X. Magnetic buoyancy and the solar dynamo. *Astrophys. J.* **198,** 205–209 (1975).
15. Brown, B. P., Browning, M. K., Brun, A. S., Miesch, M. S. & Toomre, J. Persistent magnetic wreaths in a rapidly rotating Sun. *Astrophys. J.* **711,** 424–438 (2010).
16. Dobler, W., Stix, M. & Brandenburg, A. Magnetic field generation in fully convective rotating spheres. *Astrophys. J.* **638,** 336–347 (2006).
17. Browning, M. K. Simulations of dynamo action in fully convective stars. *Astrophys. J.* **676,** 1262–1280 (2008).
18. Vilhu, O. The nature of magnetic activity in lower main sequence stars. *Astron. Astrophys.* **133,** 117–126 (1984).

19. Johns-Krull, C. M. & Valenti, J. A. Detection of strong magnetic fields on M dwarfs. *Astrophys. J.* **459,** 95–99 (1996).
20. Irwin, J. *et al.* On the angular momentum evolution of fully convective stars: rotation periods for field M-dwarfs from the MEarth transit survey. *Astrophys. J.* **727,** 56 (2011).
21. Spruit, H. C. & Ritter, H. Stellar activity and the period gap in cataclysmic variables. *Astron. Astrophys.* **124,** 267–272 (1983).
22. Garraffo, C., Drake, J. J. & Cohen, O. Magnetic complexity as an explanation for bimodal rotation populations among young stars. *Astrophys. J.* **807,** L6 (2015).
23. Nelson, N. J., Brown, B. P., Brun, A. S., Miesch, M. S. & Toomre, J. Magnetic wreaths and cycles in convective dynamos. *Astrophys. J.* **762,** 73 (2013).
24. Fan, Y. & Fang, F. A simulation of convective dynamo in the solar convective envelope: maintenance of the solar-like differential rotation and emerging flux. *Astrophys. J.* **789,** 35 (2014).
25. Dikpati, M. & Gilman, P. A. Simulating and predicting solar cycles using a flux-transport dynamo. *Astrophys. J.* **649,** 498–514 (2006).
26. Muñoz-Jaramillo, A., Nandy, D. & Martens, P. C. H. Helioseismic data inclusion in solar dynamo models. *Astrophys. J.* **698,** 461–478 (2009).
27. Cox, A. N., Hodson, S. W. & Shaviv, G. On the ratio of mixing length to scale height in red dwarfs. *Astrophys. J.* **245,** L37–L40 (1981).
28. Moss, D. L. & Taylor, R. J. The influence of a poloidal magnetic field on convection in stellar cores. *Mon. Not. R. Astron. Soc.* **147,** 133–138 (1970).
29. Mullan, D. J. & MacDonald, J. Are magnetically active low-mass M dwarfs completely convective? *Astrophys. J.* **559,** 353–371 (2001).
30. Browning, M. K., Weber, M. A., Chabrier, G. & Massey, A. P. Theoretical limits on magnetic field strengths in low-mass stars. *Astrophys. J.* **818,** 189 (2016).

Acknowledgements N.J.W. acknowledges a Royal Astronomical Society Research Fellowship and an STFC Ernest Rutherford Fellowship. J.J.D. was supported by NASA contract NAS8-03060 to the Chandra X-ray Center. We thank J. Irwin, R. Jeffries and A. West for assistance and comments on an early draft of this paper. This research has made use of the Vizier (http://vizier.u-strasbg.fr/cgi-bin/VizieR) and SIMBAD (http://simbad.u-strasbg.fr/simbad/sim-fid) databases (operated at CDS, Strasbourg, France).

Author Contributions N.J.W. reduced the Chandra observations, measured the X-ray fluxes and made the necessary calculations to plot the stars in Fig. 1. N.J.W. and J.J.D. wrote the interpretation and discussion of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.J.W. (nick.nwright@gmail.com).

Reviewer Information *Nature* thanks D. Moss and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

New observations. Two targets were chosen from the MEarth Transit survey[20] of fully convective stars: G184-31 and GJ 3253, of spectral types M4.5[31] and M5[32] with rotation periods of 83.8 days (ref. 33) and 78.8 days (ref. 20), respectively.

The targets were observed with the Advanced CCD Imaging Spectrometer (ACIS)[34] on the Chandra X-ray Observatory[35] using the ACIS-S (spectroscopic) CCD array. Observations were performed in 'very faint' mode and were placed on the back-illuminated S3 (the 3rd CCD chip in the ACIS-S spectroscopic array) chip (owing to its higher sensitivity to soft X-rays. The observations were performed on 6 December 2012 and 25 September 2013 for G184-31 and GJ 3253 respectively, with exposure times of 8.0 ks and 21.0 ks.

Observations were processed using the CIAO 4.5 software tools[36] and the CALDB 4.5.8 calibration files following standard procedures. The two sources were clearly identified and detected with significances of 5.0σ and 10.2σ at their expected positions. Point-source extraction was performed using CIAO 4.5. The total number of net counts was measured to be 35^{+7}_{-6} and 124^{+12}_{-11} for G184-31 and GJ 3253, respectively. Light curves were constructed for both sources to search for high levels of variability that might inflate the quiescent flux level measured, but no significant variability was detected.

Thermal plasma X-ray spectral models were fitted to the extracted spectra using XSPEC[37] version 12.6.0 and compared to APEC[38] single-temperature optically thin model spectra of an absorbed thermal plasma in collisional ionization equilibrium, allowing the plasma temperature ($k_B T$, where k_B is Boltzmann's constant and T is temperature) and the hydrogen column density (N_H) to vary freely. A grid of initial thermal plasma temperatures covering the range $k_B T = 0.1$–3.0 keV was used to prevent fitting to local minima. The model with the lowest Cash statistic[39] was selected as the best fit for each source (the Cash statistic is an application of the likelihood ratio test that is suitable for low-signal data). The best-fitting thermal plasma temperatures were found to be $k_B T = 0.78 \pm 0.13$ keV and 0.30 ± 0.05 keV for G184-31 and GJ 3253, respectively, consistent with the values found for other M-type dwarf stars. The fitted hydrogen column density is consistent with no absorption, as expected for the proximity of these sources.

Absorption-corrected broadband (0.5–8.0 keV) fluxes of $F_X = (2.08 \pm 0.38) \times 10^{-14}$ erg/s/cm^2 and $(4.35 \pm 0.41) \times 10^{-14}$ erg/s/cm^2 were calculated from the model fits for G184-31 and GJ 3253 respectively. Combined with their known parallax distances[31,32] the fluxes were used to calculate X-ray luminosities in the ROSAT band (0.1–2.4 keV), for consistency and ease of comparison with previous studies[7]. Fractional X-ray luminosities were then calculated from the observed J-band magnitudes and the appropriate bolometric corrections[40].

Literature data. We searched the literature for fully convective stars (spectral type M4 or later) with existing measured rotation periods and X-ray luminosities, excluding any with short rotation periods ($P_{rot} < 20$ days), which would place the object in the saturated regime of Fig. 1. Two stars were found that met our criteria: GJ 699 (Barnard's star, M4V; ref. 41) and GJ 551 (Proxima Centauri, M5.5V; ref. 42) with rotation periods[43] and X-ray luminosity values[44] existing in the literature. Rossby numbers and fractional X-ray luminosities for these stars were calculated as for the two newly observed stars.

We also uncovered a number of stars that are close to the convective boundary in the literature[45,46] (spectral types M3 or M3.5). The spectral types and colours[9] of these stars imply that they are partly convective and so they are shown in Fig. 1 as grey dots.

Sample size. No statistical methods were used to predetermine sample size.

Code availability. The CIAO code used to reduce the Chandra X-ray Observatory data are available at http://cxc.cfa.harvard.edu/ciao and the associated calibration database can be found at http://cxc.cfa.harvard.edu/caldb. The XSPEC code used to perform X-ray spectral fitting is available at https://heasarc.gsfc.nasa.gov/xanadu/xspec.

31. Reid, I. N. *et al.* Meeting the cool neighbors. VIII. A preliminary 20 parsec census from the NLTT catalogue. *Astron. J.* **128,** 463–483 (2004).
32. Newton, E. R. *et al.* Near-infrared metallicities, radial velocities, and spectral types for 447 nearby M dwarfs. *Astron. J.* **147,** 20 (2014).
33. West, A. A. *et al.* An activity-rotation relationship and kinematic analysis of nearby mid-to-late-type M dwarfs. *Astrophys. J.* **812,** 3 (2015).
34. Garmire, G. P., Bautz, M. W., Ford, P. G., Nousek, J. A. & Ricker, G. R. Jr. Advanced CCD Imaging Spectrometer (ACIS) instrument on the Chandra X-ray Observatory. *Proc. SPIE* **4851,** 28–44 (2003).
35. Weisskopf, M. C., Tananbaum, H. D., Van Speybroeck, L. P. & O'Dell, S. L. Chandra X-ray Observatory (CXO): overview. *Proc. SPIE* **4012,** 2–16 (2000).
36. Fruscione, A. *et al.* CIAO: Chandra's data analysis system. *Proc. SPIE* **6270,** 62701V (2006).
37. Arnauld, K. A. XSPEC: the first ten years. *ASP Conf. Series* **101,** 17–20 (Astronomical Society of the Pacific, 1996).
38. Smith, R. K., Brickhouse, N. S., Liedahl, D. A. & Raymond, J. C. Collisional plasma models with APEC/APED: emission-line diagnostics of hydrogen-like and helium-like ions. *Astrophys. J.* **556,** L91–L95 (2001).
39. Cash, W. Parameter estimation in astronomy through application of the likelihood ratio. *Astrophys. J.* **228,** 939–947 (1979).
40. Cruz, K. L. & Reid, I. N. Meeting the cool neighbors. III. Spectroscopy of northern NLTT stars. *Astron. J.* **123,** 2828–2840 (2002).
41. Kirkpatrick, J. D., Henry, T. J. & McCarthy, D. W. Jr. A standard stellar spectral sequence in the red/near-infrared – classes K5 to M9. *Astrophys. J. Suppl. Ser.* **77,** 417–440 (1991).
42. Bessell, M. S. The late-M dwarfs. *Astron. J.* **101,** 662–676 (1991).
43. Benedict, G. F. *et al.* Photometry of Proxima Centauri and Barnard's Star using Hubble Space Telescope Fine Guidance Sensor 3: a search for periodic variations. *Astron. J.* **116,** 429–439 (1998).
44. Schmitt, J. H. M. M. & Liefke, C. NEXXUS: a comprehensive ROSAT survey of coronal X-ray emission among nearby solar-like stars. *Astron. Astrophys.* **417,** 651–665 (2004).
45. Kiraga, M. & Stepien, K. Age-rotation-activity relations for M dwarf stars based on ASAS photometric data. *Acta Astron.* **57,** 149–168 (2007).
46. Jeffries, R. D., Jackson, R. J., Briggs, K. R., Evans, P. A. & Pye, J. P. Investigating coronal saturation and supersaturation in fast-rotating M-dwarf stars. *Mon. Not. R. Astron. Soc.* **411,** 2099–2112 (2011).

doi:10.1038/nature18960

Combinatorial design of textured mechanical metamaterials

Corentin Coulais[1,2], Eial Teomy[3], Koen de Reus[1], Yair Shokef[3] & Martin van Hecke[1,2]

The structural complexity of metamaterials is limitless, but, in practice, most designs comprise periodic architectures that lead to materials with spatially homogeneous features[1–11]. More advanced applications in soft robotics, prosthetics and wearable technology involve spatially textured mechanical functionality, which requires aperiodic architectures. However, a naive implementation of such structural complexity invariably leads to geometrical frustration (whereby local constraints cannot be satisfied everywhere), which prevents coherent operation and impedes functionality. Here we introduce a combinatorial strategy for the design of aperiodic, yet frustration-free, mechanical metamaterials that exhibit spatially textured functionalities. We implement this strategy using cubic building blocks—voxels—that deform anisotropically, a local stacking rule that allows cooperative shape changes by guaranteeing that deformed building blocks fit together as in a three-dimensional jigsaw puzzle, and three-dimensional printing. These aperiodic metamaterials exhibit long-range holographic order, whereby the two-dimensional pixelated surface texture dictates the three-dimensional interior voxel arrangement. They also act as programmable shape-shifters, morphing into spatially complex, but predictable and designable, shapes when uniaxially compressed. Finally, their mechanical response to compression by a textured surface reveals their ability to perform sensing and pattern analysis. Combinatorial design thus opens up a new avenue towards mechanical metamaterials with unusual order and machine-like functionalities.

The architecture of a material is crucial for its properties and functionality. This connection between form and function is leveraged by mechanical metamaterials[1–15], whose patterned microstructures are designed to obtain unusual behaviours such as negative response parameters[1,9], multistability[5,8,11] or programmability[8,12]. For ordinary materials, aperiodic architectures and structural complexity are associated with geometric frustration[16,17], which prevents a coherent and predictable response. Frustration hinders functionality, and metamaterial designs have therefore focused on periodic structures. However, the high level of control provided by 3D printing provokes the question of whether one can design and create structurally complex, yet frustration-free, metamaterials[18].

We investigate this question using a combinatorial design strategy. We assemble 1-cm^3 flexible, cubic building blocks, or voxels, into a cubic lattice that then forms a metamaterial (Fig. 1a). These building blocks are anisotropic and have one soft mode of deformation aligned with their internal axis of orientation, resulting in elongated or flattened shapes that we refer to as 'bricks' with positive or negative polarization (Fig. 1b). Generally, mechanical metamaterials with randomly orientated building blocks are frustrated, because it is impossible for all blocks to cooperatively deform according to their soft mode: the bricks do not fit. We call voxelated[19] metamaterials that allow soft deformations, or, equivalently, in which all bricks fit, 'compatible'. A trivial example of a compatible configuration is a periodic stacking of alternately polarized, parallelly orientated bricks. Hence, a periodic metamaterial consisting of parallel blocks is expected to exhibit a collective and harmonious deformation mode. We realized such a metamaterial by a combination of 3D printing and moulding (see Methods). Uniaxial compression indeed triggers a collective pattern change[2] in three dimensions and produces the expected staggered configuration in which each brick is adjacent to six bricks of opposite polarization (Fig. 1c; Supplementary Video 1).

We now consider the way in which to design aperiodic, yet frustration-free, mechanical metamaterials. Crucially, the internal structure of our blocks is anisotropic and each block can be oriented independently to allow structurally complex architectures. We think of these blocks as voxels, represent their orientation at each lattice point with a colour (Fig. 1d) and explore the discrete design space of such voxelated metamaterials. Compatibility requires two conditions to be met. First, pairs

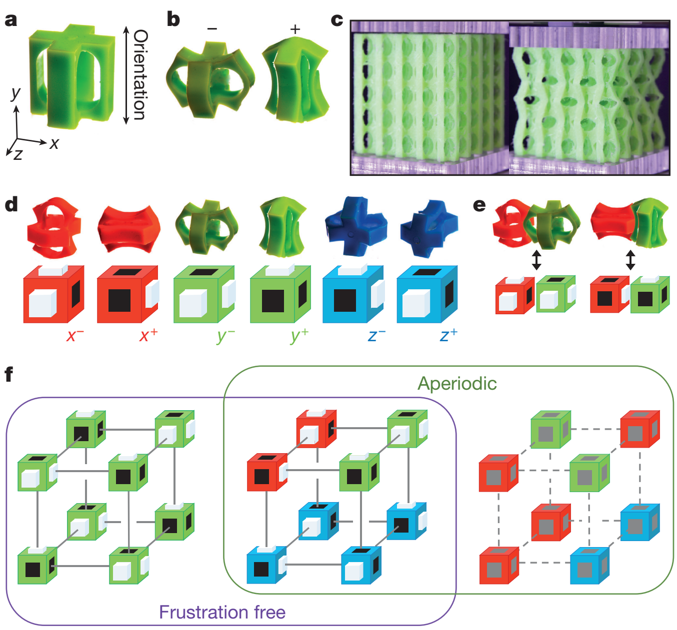


Figure 1 | Voxelated mechanical metamaterials. a, A flexible anisotropic building block in its undeformed state. **b**, Corresponding flattened (−) and elongated (+) deformed bricks. **c**, A $5 \times 5 \times 5$ metacube consisting of parallel blocks (left) shows a collective deformation under uniaxial compression (right). **d**, Bricks (top) and their schematic representation (bottom), with the colour indicating orientation, and black dents and white protrusions representing deformations. **e**, Two examples where appropriately polarized, adjacent bricks fit together. **f**, Periodic (left), complex (middle) and frustrated (right) $2 \times 2 \times 2$ stackings—for the frustrated stacking, no consistent brick configurations exist (grey/dashed). Schematic symbols are separated from one another for ease of visualization.

[1]Huygens-Kamerlingh Onnes Laboratory, Universiteit Leiden, PO box 9504, 2300 RA Leiden, The Netherlands. [2]FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands. [3]School of Mechanical Engineering and The Sackler Center for Computational Molecular and Materials Science, Tel Aviv University, Tel Aviv 69978, Israel.

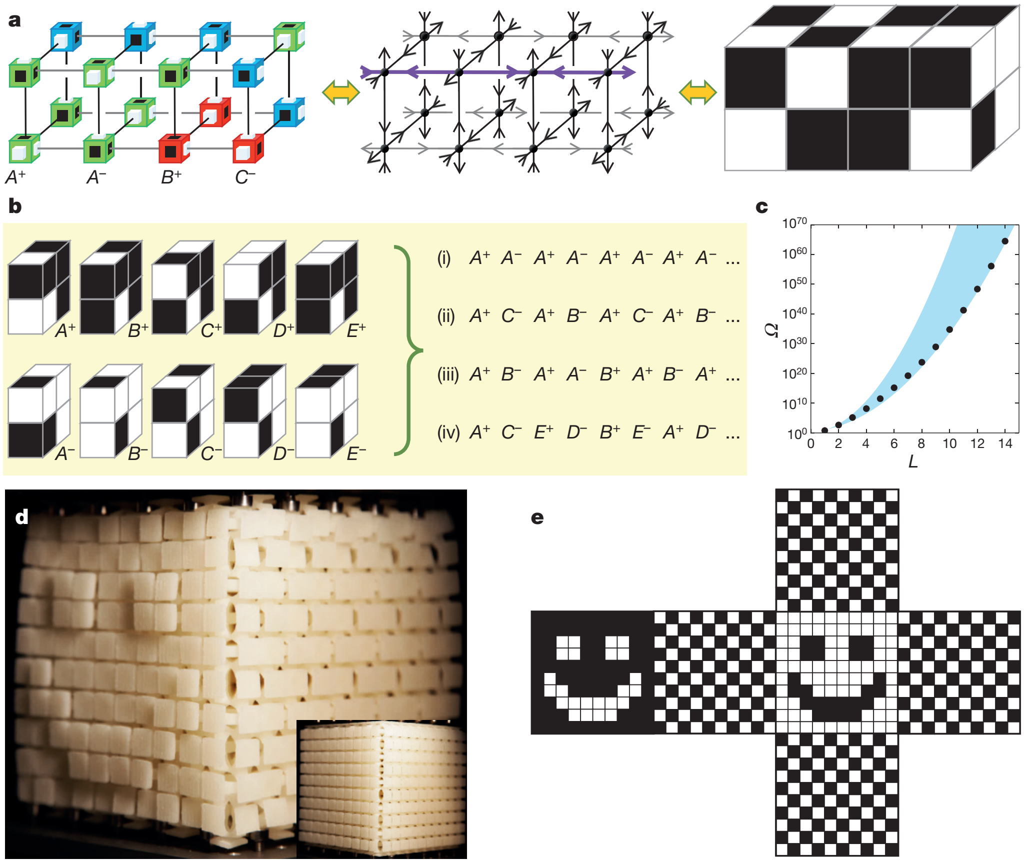


Figure 2 | Combinatorial design. a, Mapping of bricks to internal spins and then to surface spins (from left to right). **b**, Set of five pairs of motifs compatible with a given x texture, and examples of periodic (i, ii), quasi-periodic (iii; beginning of the Fibonacci sequence) and aperiodic (iv) motif stackings. **c**, Exact number of compatible $L \times L \times L$ spin configurations Ω (black circles) and lower and upper bounds (blue shaded region); see Methods and Supplementary Information. **d**, A $10 \times 10 \times 10$ metacube reveals its precisely designed surface texture under uniaxial compression. Square surface pedestals added for visualization. Inset, the undeformed metacube. See Methods for experimental details and Supplementary Videos 2–4. **e**, Schematic representation of the deformations at all surfaces of the metacube in **d**.

of neighbouring bricks should exhibit closely matching shapes along their common face. Our building blocks are precisely designed such that, given the polarization of one brick, the polarization of an adjacent brick can be adapted so that the pair have a tight fit, irrespective of their mutual orientation. Hence, we need to track only the outward or inward deformations of the surfaces of the building blocks (Fig. 1e). The second compatibility condition concerns the combinatorics of the voxel arrangement: all bricks should fit, such that protrusions and depressions of all neighbouring bricks are matched. In general, the first condition can be met by clever building-block design, whereas the second condition leads to a complex combinatorial 3D tiling problem.

As we will show, although the compatibility condition is violated in most random configurations, which are consequently frustrated, our specific building blocks allow for many complex configurations where all protrusions and depressions match. These non-parallel, structurally complex, yet compatible, architectures combine the rich spatial texture of aperiodic materials with the predictability of ordered materials, and form the blueprint for aperiodic, frustration-free mechanical metamaterials (Fig. 1f).

The design of complex architectures is simplified by mapping brick configurations to spin configurations that satisfy a so-called 'ice rule'[20–23] and, as such, is reminiscent of tiling[24] and constraint-satisfaction[25,26] problems. We identify each brick with a vertex, connected to neighbouring vertices by bonds that represent the common face between bricks (Fig. 2a). Dents and bumps map to inward or outward spins σ_x, σ_y, σ_z, and by placing a single spin per bond, the first compatibility condition is trivially satisfied. The second condition maps to the ice rule, which stipulates that the six bonds of each vertex should correspond to a brick configuration, whereby the six bricks, x^-, x^+, y^-, y^+, z^- and z^+, correspond to spin configurations $(\sigma_x, \sigma_y, \sigma_z) = (-++)$, $(+--)$, $(+-+)$, $(-+-)$, $(++-)$ and $(--+)$, respectively. Evidently, each allowed spin configuration corresponds to a compatible brick stacking and corresponding voxel configuration. Conversely, each compatible voxel configuration corresponds to two spin configurations related by parity (spin flip)—a symmetry that originates from the opposite polarizations allowed by each building block.

All compatible metamaterials thus obtained feature an unusual form of long-range order that relates the boundary to the bulk. Owing to the reflection symmetry of the bricks, spins along lines of bonds are alternating. Therefore, spins at opposing boundaries are equal (opposite) when their distance is odd (even). Because spins at the surface of a metamaterial represent its texture of bumps and dents, this property implies that textures at opposite faces of a metacube are directly linked. Moreover, once the surface texture is fixed, all internal spins and therefore bricks are determined (Fig. 2a). We call this unusual relation between surface and bulk 'holographic order' (see Methods).

The combination of parity and holographic order implies that any motif consisting of $n \times p \times q$ compatible blocks can be stacked in a space-filling manner, because the surface spins of such motifs have compatible textures. Moreover, once the x spins are fixed along a plane, we can determine a set containing all motifs A^+, B^+, … with matching x spins and by parity obtain A^-, B^-, … (Fig. 2b and Methods). These motifs can be stacked in an arbitrary order, as long as we alternate between '+' and '−' motifs; this allows the straightforward design of periodic, quasi-periodic and aperiodic metamaterials (Fig. 2b).

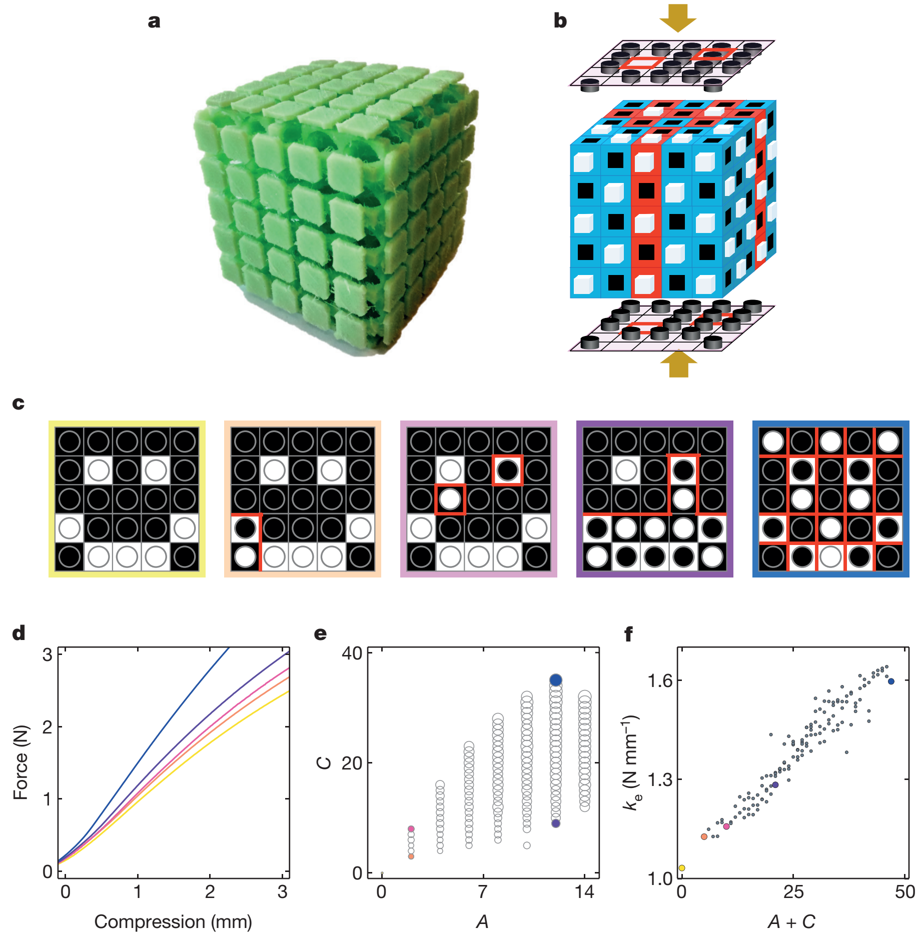


Figure 3 | Pattern recognition and pattern analysis. a, Experimental realization of an elastic $5 \times 5 \times 5$ metacube programmed for a 'smiley' texture. **b**, Schematic of experiments where the metacube in **a** is compressed between patterned clamps. The red and blue building blocks are oriented as in Fig 1; cylinders signify stubs that are positioned at sites where $\sigma_z^K = 1$ and red lines demarcate regions where $\sigma_z^K = \sigma_z^L$ and $\sigma_z^K = -\sigma_z^L$. **c**, Examples of mismatch between cube lock texture $\{\sigma_z^L\}$ (squares) and boundary key textures $\{\sigma_z^K\}$ (circles). Red lines as in **b**. **d**, Experimental force–compression curves. Colour corresponds to the key textures in **c**. **e**, The experimentally obtained stiffness k_e varies systematically with both the area (A) and circumference (C) of the mismatch. Coloured data points correspond to the key textures in **c**, and the size of each circle represents k_e (the yellow data point at $A = C = 0$ corresponds to very small k_e and so is hard to see). **f**, The stiffness k_e is essentially linear in $A + C$. Coloured data points as in **e**; small filled grey circles correspond to the open circles in **e**.

By removing building blocks at the boundary, complex shapes can be realized; but for simplicity we focus here on cubic metamaterials.

Holographic order substantially restricts the number of potential compatible configurations: whereas for generic configurations their multitude is set by the volume, for compatible configurations it is set by the surface area. Moreover, many surface textures lead to forbidden internal vertices such as where all spins are equal. For example, in general it is not possible to arbitrarily choose the surface texture at two faces simultaneously. Nevertheless, the number of distinct $L \times L \times L$ (equivalent to $n \times p \times q$, above, with $n = p = q := L$) spin configurations $\Omega(L)$ is extremely large. To quantify the design limits and possibilities, we exactly evaluated $\Omega(L)$ up to $L = 14$ where $\Omega \approx 3 \times 10^{64}$, and obtained strict and asymptotic lower and upper bounds (Fig. 2c, Methods and Supplementary Information).

Despite the limitations imposed by compatibility, the design space of voxelated metamaterials is huge. To illustrate this, we constructed a general algorithm to obtain all $L \times L \times 1$ motifs that are compatible with a given texture $\{\sigma_z\}$ (see Methods and Supplementary Information); for each texture there are at least two distinct motifs. We now show that we can use this algorithm to design arbitrarily pixelated patterns of bumps and depressions, or textures, at a given surface of a metacube as a step towards arbitrary shape-morphing materials. In Fig. 2d we show a rationally designed metacube created by 3D printing. Under uniaxial compression (see Methods), the initially flat surface of this cube reveals its spatial texture, with the front and back related by holography (Fig. 2e, Methods and Supplementary Videos 2–4). This cooperative, complex, yet controlled, shape morphing illustrates that our combinatorial method allows for the rational design of shape-shifting metamaterials.

We finally show that when aperiodic metacubes are compressed by patterned surfaces their response can be used for mechanical pattern analysis. We created a compatible $5 \times 5 \times 5$ metacube, programmed with a 'smiley' texture $\pm\{\sigma_z^L\}$ that acts as a 'lock' (see Methods), which is compressed between two identical surfaces that have a pixelated 'key' texture $\{\sigma_z^K\}$, created by placing 18 stubs in templated clamps (Fig. 3a, b). We characterize the difference between lock and key patterns by the area or number of misplaced stubs, A, as well as the circumference of the misplaced area, C (Fig. 3c), and use 136 different key patterns, which cover all possible values of A and C that can be reached by 18 stubs. For each key, we performed experiments (simulations) to determine the stiffness k_e (k_s) via the slope of the force–compression curves—the values agree very closely with one another (see Methods). When the key equals one of the two lock textures $\pm\{\sigma_z^L\}$, all bricks deform compatibly and the stiffness k is low. Incompatibly textured surfaces push metacubes into frustrated states, leading to an energy penalty and increased stiffness (Fig. 3d). The increase with A evidences

simple lock-and-key functionality, but when k is plotted as function of A and C, it is seen to increase with C also—for the same number of misplaced stubs, a range of stiffnesses can be observed (Fig. 3e). When plotted as function of $A+C$, all our data collapse onto a straight line, which evidences intricate collective phenomena (Fig. 3f). We suggest that, owing to parity, different parts of the cube deform in opposing parity, and that the stiffness is determined by the size of the domain walls separating these regions, which is given by $A+C$ (see Methods). Together, these findings demonstrate the ability of a metacube to perform an arithmetical calculation on the mismatch between key and lock patterns, in behaviour more readily associated with machines than with materials.

Combinatorial strategies enable the design of machine-like materials that can be programmed with specific shape-sensing and shape-shifting tasks. We anticipate that combinatorial design of textured metamaterials can be extended in various directions. First, the inclusion of vacancies could lead to multishape materials[27], whereas defects can induce controlled frustration to obtain multistability, memory and programmability[8,11,12]. Second, differently shaped building blocks, such as triangles or hexagons in two dimensions and truncated octahedra or gyrobifastigii in three dimensions, or mixtures of building blocks could be used to tile space. Third, building blocks with degrees of freedom that are different from the simple 'inwards or outwards' deformations considered here could be considered—a prime example being origami units that have folding motions[10–12]. Finally, heating or magnetic fields instead of compression could be used to actuate shape-changing metamaterials, while non-mechanical textured functionalities such as wavefront shaping could also be achieved. We envision a range of applications for which control and processing of spatially complex mechanical information is key. Textured metamaterials can be designed to naturally interface with the complex shapes and shapeability of the human body, in prosthetics, haptic devices and wearables. Moreover, shape changing is central to a wide variety of actuators and sensors, in particular in the context of soft robots[28–30]. Finally, at smaller scales, controllable surface textures could control friction, wetting and drag.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 March; accepted 17 June 2016.

1. Lakes, R. Foam structures with a negative Poisson's ratio. *Science* **235,** 1038–1040 (1987).
2. Mullin, T., Deschanel, S., Bertoldi, K. & Boyce, M. C. Pattern transformation triggered by deformation. *Phys. Rev. Lett.* **99,** 084301 (2007).
3. Grima, J. N. & Evans, K. E. Auxetic behavior from rotating squares. *J. Mater. Sci. Lett.* **19,** 1563–1565 (2000).
4. Schaedler, T. A. *et al.* Ultralight metallic microlattices. *Science* **334,** 962–965 (2011).
5. Nicolaou, Z. G. & Motter, A. E. Mechanical metamaterials with negative compressibility transitions. *Nat. Mater.* **11,** 608–613 (2012).
6. Babaee, S., Shim, J., Weaver, J. C., Patel, N. & Bertoldi, K. 3D soft metamaterials with negative Poisson's ratio. *Adv. Mater.* **25,** 5044–5049 (2013).
7. Kadic, M., Bückmann, T., Schittny, R. & Wegener, M. Metamaterials beyond electromagnetism. *Rep. Prog. Phys.* **76,** 126501 (2013).
8. Florijn, B., Coulais, C. & van Hecke, M. Programmable mechanical metamaterials. *Phys. Rev. Lett.* **113,** 175503 (2014).
9. Coulais, C., Overvelde, J. T. B., Lubbers, L. A., Bertoldi, K. & van Hecke, M. Discontinuous buckling of wide beams and metabeams. *Phys. Rev. Lett.* **115,** 044301 (2015).
10. Schenk, M. & Guest, S. D. Geometry of Miura-folded metamaterials. *Proc. Natl Acad. Sci. USA* **110,** 3276–3281 (2013).
11. Waitukaitis, S., Menaut, R., Chen, B. G. & van Hecke, M. Origami multistability: from single vertices to metasheets. *Phys. Rev. Lett.* **114,** 055503 (2015).
12. Silverberg, J. L. *et al.* Using origami design principles to fold reprogrammable mechanical metamaterials. *Science* **345,** 647–650 (2014).
13. Bückmann, T., Thiel, M., Kadic, M., Schittny, R. & Wegener, M. An elasto-mechanical unfeelability cloak made of pentamode metamaterials. *Nat. Commun.* **5,** 4130 (2014).
14. Paulose, J., Meeussen, A. S. & Vitelli, V. Selective buckling via states of self-stress in topological metamaterials. *Proc. Natl Acad. Sci. USA* **112,** 7639–7644 (2015).
15. Liu, Z. *et al.* Locally resonant sonic materials. *Science* **289,** 1734–1736 (2000).
16. Wannier, G. H. Antiferromagnetism: the triangular Ising net. *Phys. Rev.* **79,** 357–364 (1950).
17. Sadoc, J. F. & Mosseri, R. *Geometrical Frustration* (Cambridge Univ. Press, 1999).
18. Zykov, V., Mytilinaios, E., Adams, B. & Lipson, H. Robotics: Self-reproducing machines. *Nature* **435,** 163–164 (2005).
19. Ware, T. H., McConney, M. E., Wie, J. J., Tondiglia, V. P. & White, T. J. Voxelated liquid crystal elastomers. *Science* **347,** 982–984 (2015).
20. Harris, M. J., Bramwell, S. T., McMorrow, D. F., Zeiske, T. & Godfrey, K. W. Geometrical frustration in the ferromagnetic pyrochlore $Ho_2Ti_2O_7$. *Phys. Rev. Lett.* **79,** 2554–2557 (1997).
21. Wang, R. F. *et al.* Artificial 'spin ice' in a geometrically frustrated lattice of nanoscale ferromagnetic islands. *Nature* **439,** 303–306 (2006).
22. Castelnovo, C., Moessner, R. & Sondhi, S. L. Magnetic monopoles in spin ice. *Nature* **451,** 42–45 (2008).
23. Nisoli, C., Moessner, R. & Schiffer, P. *Colloquium*: Artificial spin ice: designing and imaging magnetic frustration. *Rev. Mod. Phys.* **85,** 1473–1490 (2013).
24. Grünbaum, B. & Shephard, G. C. *Tilings and Patterns* (Freeman, 1987).
25. Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P. Optimization by simulated annealing. *Science* **220,** 671–680 (1983).
26. Mezard, M., Parisi, G. & Virasoro, M. A. *Spin Glass Theory and Beyond* (World Scientific, 1987).
27. Cho, Y. *et al.* Engineering the shape and structure of materials by fractal cut. *Proc. Natl Acad. Sci. USA* **111,** 17390–17395 (2014).
28. Leong, T. G. *et al.* Tetherless thermobiochemically actuated microgrippers. *Proc. Natl Acad. Sci. USA* **106,** 703–708 (2009).
29. Shepherd, R. F. *et al.* Multigait soft robot. *Proc. Natl Acad. Sci. USA* **108,** 20400–20403 (2011).
30. Overvelde, J. T., Kloek, T., Dhaen, J. J. & Bertoldi, K. Amplifying the response of soft actuators by harnessing snap-through instabilities. *Proc. Natl Acad. Sci. USA* **112,** 10863–10868 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to J. Mesman for technical support. We thank R. Golkov, Y. Kamir, G. Kosa, K. Kuipers, F. Leoni, W. Noorduin and V. Vitelli for discussions. We acknowledge funding from the Netherlands Organisation for Scientific Research through grants VICI No. NWO-680-47-609 (M.v.H. and C.C.) and VENI No. NWO-680-47-445 (C.C.) and from the Israel Science Foundation through grant numbers 617/12 and 1730/12 (E.T and Y.S.).

Author Contributions C.C. and M.v.H. conceived the main concepts. C.C., E.T., Y.S. and M.v.H. formulated the spin problem. E.T. and Y.S. solved the spin problem. C.C and K.d.R. performed the experiments and simulations with inputs from E.T., Y.S. and M.v.H. C.C. and M.v.H wrote the manuscript with contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.C. (coulais@amolf.nl).

METHODS

Combinatorial design. The presence of holographic order reduces the number of potential compatible $L \times L \times L$ spin configurations from $2 \times 3^{L^3}$ (L^3 blocks with 3 orientations and parity) to 2^{3L^2} (3 pairs of opposing surfaces each with L^2 spins). The vast majority of these are not compatible and to better understand our design space, we discuss how to exactly evaluate and obtain upper and lower bounds for $\Omega(L)$—the number potentially compatible $L \times L \times L$ spin configurations. We construct cubes by stacking motifs, and Q counts the number of $L \times L \times 1$ motifs compatible with a given texture $\{\sigma_z\}$.

To understand the possible motifs, we now classify the patterns of z bricks that arise in $L \times L \times 1$ motifs. Crucially, z bricks are sources or sinks for the in-plane spins σ_x and σ_y, and therefore each 2×2 submotif can contain only 0, 2 or 4 z bricks. This restricts the patterns of z bricks to columns, rows, and intersecting columns and rows. In general, we can enumerate the patterns of z bricks using binary column and row vectors with elements c_i and r_j, respectively, and placing a z brick at location (i, j) only when $c_i \neq r_j$ (Extended Data Fig. 1a). On the z-bricks, the z-spins form checkerboard patterns, whereas the spins in the remainder of the pattern can be chosen at will by filling each position with either an x or y brick.

In the absence of z bricks, we can obtain two motifs by fixing σ_x and σ_y to be opposite and alternating; that is, at site (i, j) of the motif, $\sigma_x = -\sigma_y = (-1)^{i+j}$ or $\sigma_x = -\sigma_y = (-1)^{i+j+1}$. In both cases, each vertex is compatible with either an x or a y brick, with corresponding positive or negative σ_z. This allows the straightforward design of two $L \times L \times 1$ motifs consistent with any $\{\sigma_z\}$. In Extended Data Fig. 1b we show these motifs, as well as four more that are compatible with a 5×5 smiley texture; hence, $Q = 6$ for this texture. In principle, $\Omega(L)$ can be exactly evaluated by determining Q for each texture $\{\sigma_z\}$, and then summing over all textures (see Supplementary Information and Extended Data Fig. 1c): $\Omega(L) := \sum_{\sigma_z} Q^L$.

Lower bound. A lower bound for Ω follows from our construction to create two motifs for any spin configuration, which implies that $Q \geq 2$. As these can be stacked in an arbitrary order, this yields at least 2^L spin configurations for a given texture. Because there are 2^{L^2} σ_z textures, we find that $\Omega(L) \geq 2^{L^2+L}$.

Staggered spins. To simplify the counting of the number of compatible spin configurations $\Omega(L)$ (for details see Supplementary Information), we define staggered spins $\tilde{\sigma}$, such that for site (i, j, k) in the metacube $\tilde{\sigma}_\alpha^{(i,j,k)} = (-1)^{i+j+k}\sigma_\alpha^{(i,j,k)}$, for $\alpha = x, y, z$. Under this invertible transformation, a checkerboard texture of $\{\sigma_z\}$, for example, corresponds to a homogeneous texture of $\{\tilde{\sigma}_z\}$ where all staggered spins are equal to either $+1$ or -1. Moreover, all sites in a given row, column or tube have the same value of $\tilde{\sigma}_x$, $\tilde{\sigma}_y$ or $\tilde{\sigma}_z$, respectively.

Upper bound. For a simple upper bound we note that the maximum value of Q is obtained when $\tilde{\sigma}_z \equiv +1$ or $\tilde{\sigma}_z \equiv -1$. For each of these textures there are $Q = 2^{L+1} - 1$ spin configurations. Consider, for example, $\tilde{\sigma}_z \equiv +1$. If $\tilde{\sigma}_x \equiv -1$, then all the $\tilde{\sigma}_y$ are free, leading to 2^L compatible structures; if $\tilde{\sigma}_y \equiv -1$, then all the $\tilde{\sigma}_x$ are free, leading to an additional 2^L compatible structures. Because $\tilde{\sigma}_x \equiv \tilde{\sigma}_y \equiv -1$ was counted twice, the total number is $2^L + 2^L - 1 = 2^{L+1} - 1$. Hence, we obtain as an upper bound: $\Omega(L) \leq 2^{L^2}(2^{L+1} - 1)^L < 2^{2L^2+L}$. A stricter upper bound is derived in Supplementary Information: $\Omega(L) \leq 4L \times (3/4)^L \times 2^{2L^2}$. The exact evaluation of $\Omega(L)$ is explained in Supplementary Information and the results are given in Extended Data Table 1.

Design limits. We note here that if n is the number of pairs of opposing surfaces where the spins can be chosen freely, then $\Omega(L) \approx 2^{nL^2}$, and that the simple lower and upper bounds given above roughly correspond to $n = 1$ and $n = 2$. Approximate calculations detailed in Supplementary Information lead to $2^{L^2+L+\log_2(3)} \leq \Omega \leq 2^{L^2+2L+2}$, and our exact evaluation of $\Omega(L)$ shows that, for large L, Ω is quite close to the lower bound (Fig. 2a). Hence, once the texture of one surface is fixed, there is limited freedom, apart from the stacking order of motifs, to design textures at other surfaces. For most spin textures, only the two simple motifs are compatible and z bricks play a minor role.

Materials and fabrication. We created the $5 \times 5 \times 5$ specimens by 3D printing water-soluble moulds, in which we cast a well-calibrated silicon rubber (PolyvinylSiloxane, Elite Double 32, Zhermarck, Young's modulus, $E = 1.32$ MPa; Poisson's ratio, $\nu \approx 0.5$). The unit bricks measure 11.46 mm × 11.46 mm × 11.46 mm, with a spherical pore of diameter $D = 10.92$ mm in the centre and four cuboid inclusions of dimension 4.20 mm × 4.20 mm × 11.46 mm at the x and y corners; see Extended Data Fig. 2a. They are stacked with a pitch of $a = 11.46$ mm, such that the filaments between the unit cells have a non-homogeneous cross-section with a minimal width $d = 0.54$ mm and a depth $w = 3.6$ mm; see Extended Data Fig. 2b.

The $10 \times 10 \times 10$ sample has the same brick dimensions and was 3D printed commercially (Materialise, Leuven, Belgium) out of sintered polyurethane ($E \approx 14$ MPa). On the faces of the aperiodic cubes, square pedestals were added to facilitate visualization of the surface texture and compression by textured boundaries.

Mechanical tests. We compressed both metacubes at a rate of 0.02 mm s^{-1} in a uniaxial testing device (Instron type 5965), which controls the compressive displacement u to better than 10 μm and measures the compressive force with an accuracy of 10^{-4} N at an acquisition rate of 0.5 Hz.

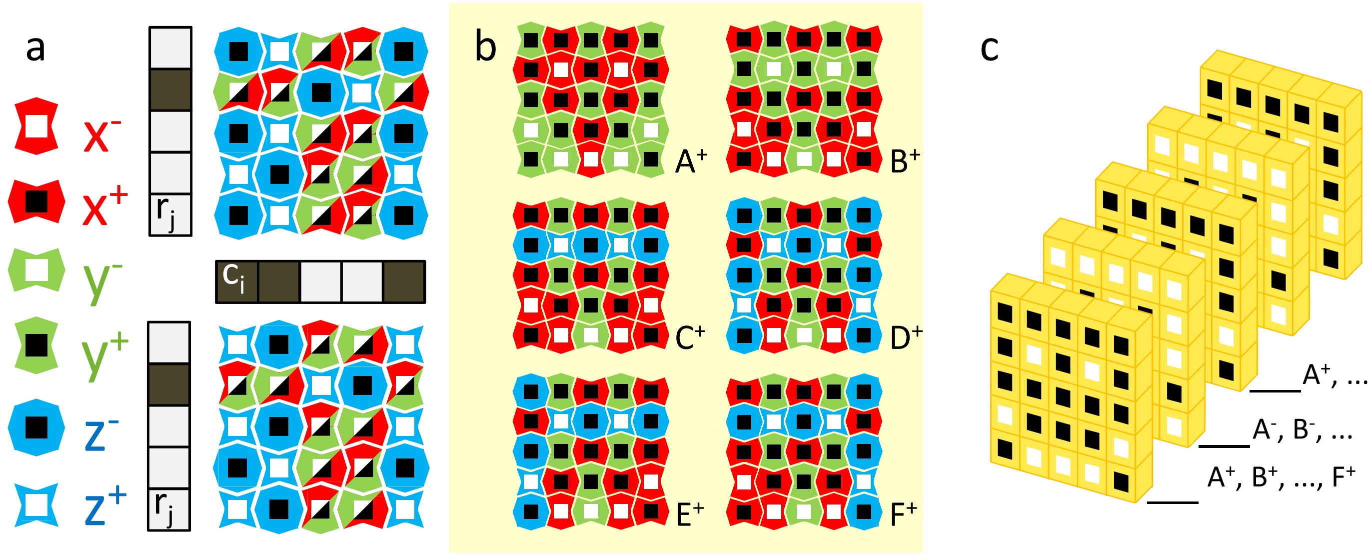
Whereas we used flat plates for Fig. 1c, we used textured boundary conditions for Figs 2d and 3. We created textured top and bottom boundaries using aluminium plates with female 3-mm threads positioned in a square array of pitch $a = 11.46 \pm 0.02$ mm, in which we mount stainless steel M3 screws whose caps were machined to a height of 3.50 ± 0.01 mm (see Extended Data Fig. 3a); this ensures precise levelling of the pins and flexibility in texture. At the start of each experiment, the cubes were carefully positioned and aligned by hand within a 1-mm accuracy on the bottom boundary. The screws were placed to form identical (complementary) top and bottom patterns for the $5 \times 5 \times 5$ ($10 \times 10 \times 10$) cube.

For the $10 \times 10 \times 10$ cube, designed as in Extended Data Fig. 4a, we used checkerboard textures that led to the desired pattern on one face (Fig. 2), the reverted pattern on the opposite face (Fig. 2 and Extended Data Fig. 4b) and checkerboard textures on the other faces (Fig. 2 and Extended Data Fig. 4c).

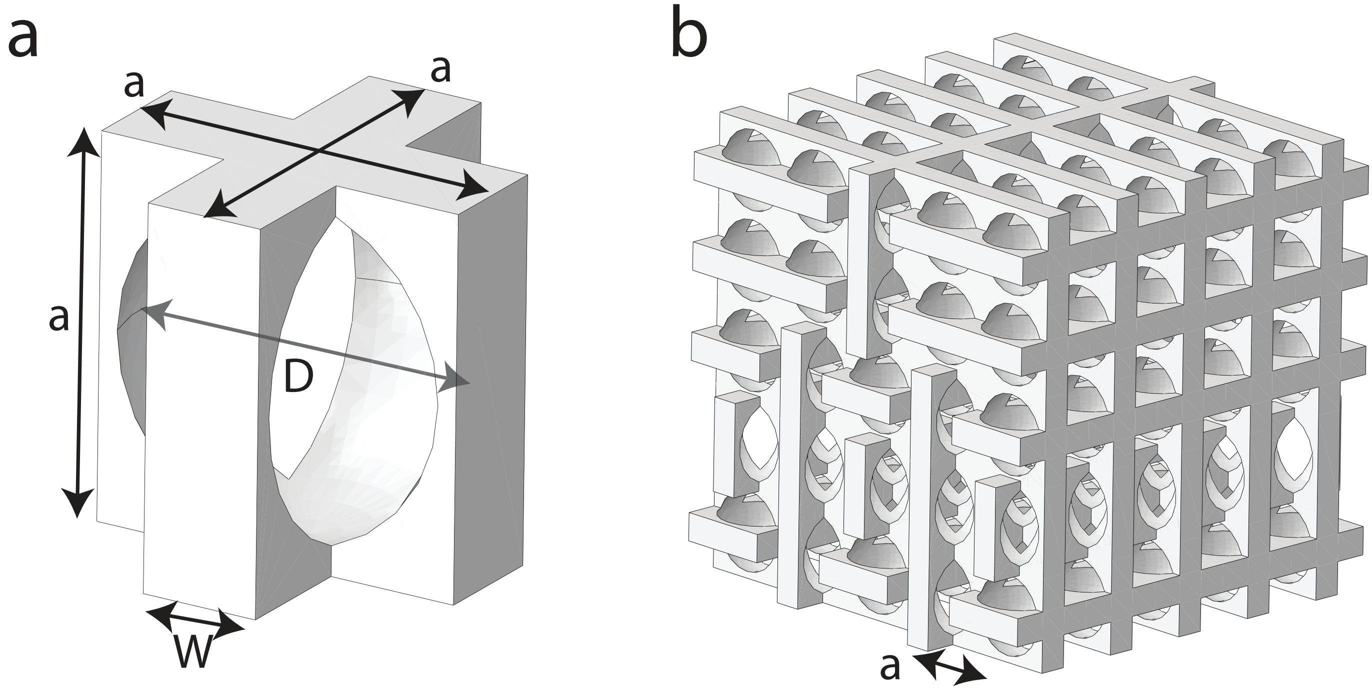
Numerical simulations. We probed the response of a $5 \times 5 \times 5$ aperiodic smiley metacube to different textures by performing a fully nonlinear analysis within the commercial package Abaqus/Standard. We modelled the elastomer using a neo-Hookean strain energy density with a Young's modulus of $E = 1.32$ MPa and a Poisson's ratio of $\nu = 0.4999$. We carried out a mesh optimization and a mesh density study leading to a typical mesh size of 0.6 mm and a total number of 1.5×10^6 hybrid quadratic tetrahedral elements (Abaqus type C3D10H). We applied uniaxial compression by applying 10 steps of magnitude $\Delta u = 0.25$ mm, using exactly the same boundary conditions and dimensions as in the experiments (Extended Data Fig. 5b).

Determination of *k*. The numerical force–displacement (F–u) curves are very well fitted by the quadratic form $F(u) = ku + \eta u^2$, which captures the effect of the nonlinear softening η and which allows an accurate estimation of the stiffness k. The experimental determination of the stiffness required special care, because small gravity-induced sagging of the cube causes a soft knee in the force–displacement curve when the top boundary makes contact with the sample. Therefore, we determined the stiffness by fitting the force–displacement curves to the same quadratic function as for the numerics, focusing on intermediate displacements $0.8\,\text{mm} \leq u \leq 2.5\,\text{mm}$ away from the knee where dF/du is linear in u; see Extended Data Fig. 5a.

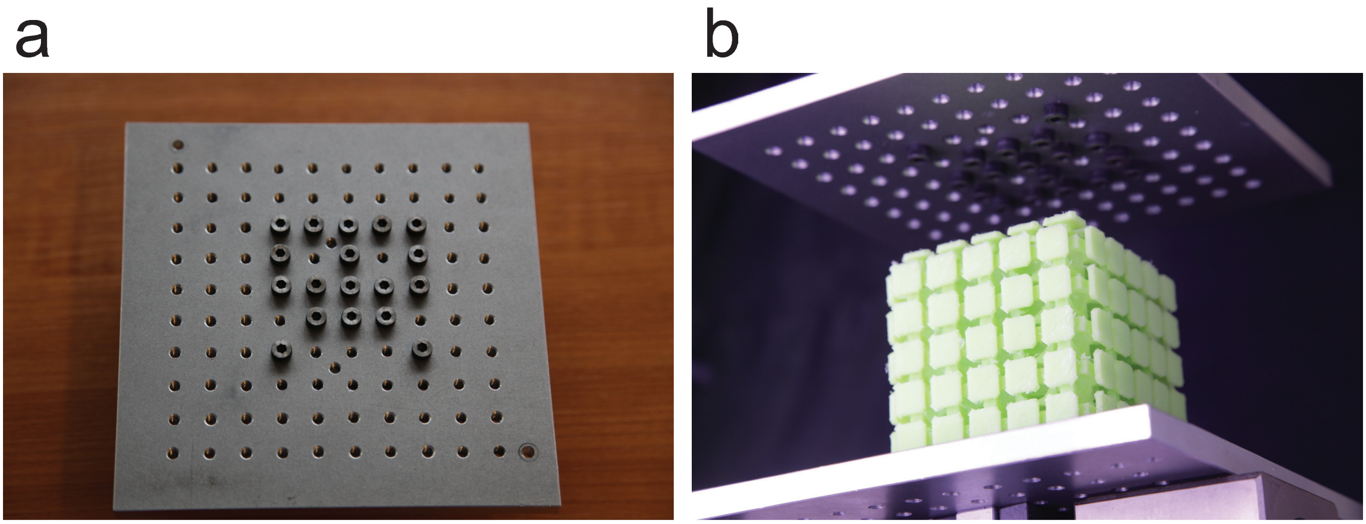
Lock and key mechanics. For the lock-and-key experiments, we used a $5 \times 5 \times 5$ cube made by stacking five $B^{\pm}$ configurations (Extended Data Fig. 1b). The key patterns consisted of 18 screws placed in a 5×5 array (see Fig. 3b, c and Extended Data Fig. 3a, b). We focused on 136 configurations that have distinct values of the area A and circumference C of the texture mismatch. For experiments and simulations, we estimated the stiffness k and observe that it increases with the mismatch between lock and key. The variation of k in experiments and simulations closely match (Extended Data Fig. 5c). Neither A nor C are good predictors for the level of frustration. To interpret the outcome of the experiments with the $5 \times 5 \times 5$ cube, we posit that, for incompatible key textures, different parts of the cube approach compatible configurations with opposite parity, thus localizing the frustration along internal domain walls. Hence, a single misplaced pixel carries an energy penalty due to the four frustrated x and y sides of the brick in front of the defect, and one z side opposing the defect—when defects touch, their interface is not frustrated. Therefore, the size of the domain walls is equal to the number of frustrated sides, which is $A + C$.



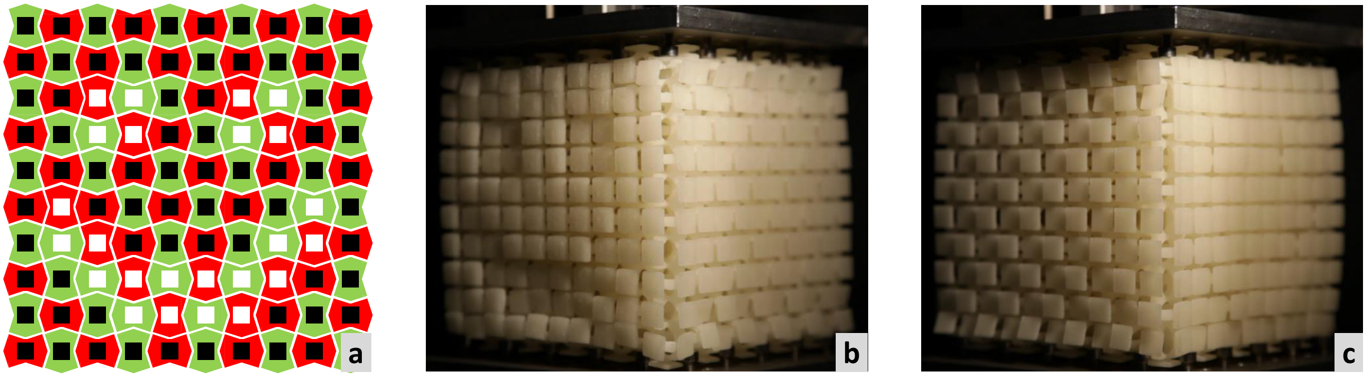
Extended Data Figure 1 | Motif-based design. a, 2D representation of the six bricks (x^-, x^+, y^-, y^+, z^-, z^+), and illustration of complex motifs. All complex motifs can be generated by defining two binary vectors with elements c_i (column) and r_j (row) that govern the placement of z bricks at location (i, j). The remaining sites are then filled with x and y bricks. Respecting parity, this generates all motifs for given $\boldsymbol{c}$ and $\boldsymbol{r}$. **b**, The six motifs (A^+, …, F^+) that are compatible with a 5×5 smiley texture. **c**, A total of 6^5 smiley metacubes can be designed by varying the stacking order; here A^- denotes the same motif as A^+ but with inversed spins. The x and y spins follow from the choice of motifs.



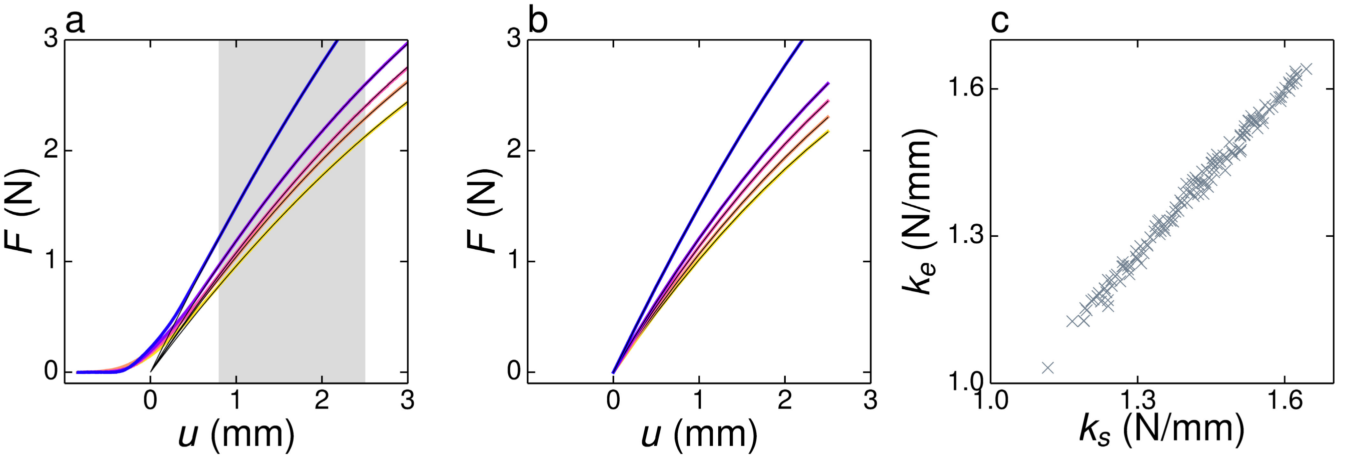
Extended Data Figure 2 | Implementation. a, **b**, Computer assisted design of the geometry of the unit cell (**a**) and a 5 × 5 × 5 cube (**b**). All our samples were 3D printed with the dimensions $a = 11.46$ mm, $D = 10.92$ mm and $w = 3.6$ mm. To make the wall thickness outside the cube equal to the internal wall thickness, the outer walls are thickened by 0.27 mm.



Extended Data Figure 3 | Lock-and-key experiment. a, Picture of the textured clamp. **b**, Side view of the experiment.



Extended Data Figure 4 | 10 × 10 × 10 metacube under uniaxial compression. a, Motif A^+—the cube is designed by stacking motifs A^+ and A^-. **b**, Opposite face of the one shown in Fig. 1e showing the inverted pattern. **c**, One of the transverse faces showing a checkerboard pattern.



Extended Data Figure 5 | Complex sensory properties of a complex $5 \times 5 \times 5$ metacube with internal smiley texture. **a**, Force–compression (F–u) curves for five experiments (thick solid lines), where the colour indicates the external texture shown in Fig. 3c. The black thin lines show fits with a quadratic function $F(u) = ku + \eta u^2$ performed in the shaded region $0.8\,\text{mm} \leq u \leq 2.5\,\text{mm}$. **b**, Corresponding numerical results. **c**, Scatter plot showing very good correspondence between the stiffness obtained from simulations (k_s) and experiments (k_e).

Extended Data Table 1 | The exact value of Ω for $L \times L \times L$ metacubes up to $L = 14$

L	Ω
1	6
2	450
3	151,206
4	145,456,074
5	325,148,366,166
6	1,562,036,085,226,890
7	17,234,732,991,509,112,246
8	578,304,084,367,752,824,053,674
9	84,438,573,424,284,282,414,882,546,966
10	58,592,971,553,875,504,020,753,814,442,326,410
11	181,442,224,689,012,470,542,563,031,429,841,423,983,926
12	2,404,888,026,041,008,595,056,652,999,310,606,919,098,996,796,074
13	126,725,905,761,644,879,286,362,510,660,061,876,041,719,518,257,045,613,846
14	30,625,852,190,216,495,511,364,347,343,665,021,261,262,812,628,299,779,541,749,100,810

LETTER

doi:10.1038/nature18640

Western boundary currents regulated by interaction between ocean eddies and the atmosphere

Xiaohui Ma[1,2], Zhao Jing[1,2], Ping Chang[1,2,3], Xue Liu[1,2], Raffaele Montuoro[3], R. Justin Small[4], Frank O. Bryan[4], Richard J. Greatbatch[5,6], Peter Brandt[5,6], Dexing Wu[1], Xiaopei Lin[1] & Lixin Wu[1]

Current climate models systematically underestimate the strength of oceanic fronts associated with strong western boundary currents, such as the Kuroshio and Gulf Stream Extensions, and have difficulty simulating their positions at the mid-latitude ocean's western boundaries[1]. Even with an enhanced grid resolution to resolve ocean mesoscale eddies—energetic circulations with horizontal scales of about a hundred kilometres that strongly interact with the fronts and currents—the bias problem can still persist[2]; to improve climate models we need a better understanding of the dynamics governing these oceanic frontal regimes. Yet prevailing theories about the western boundary fronts are based on ocean internal dynamics without taking into consideration the intense air–sea feedbacks in these oceanic frontal regions. Here, by focusing on the Kuroshio Extension Jet east of Japan as the direct continuation of the Kuroshio, we show that feedback between ocean mesoscale eddies and the atmosphere (OME-A) is fundamental to the dynamics and control of these energetic currents. Suppressing OME-A feedback in eddy-resolving coupled climate model simulations results in a 20–40 per cent weakening in the Kuroshio Extension Jet. This is because OME-A feedback dominates eddy potential energy destruction, which dissipates more than 70 per cent of the eddy potential energy extracted from the Kuroshio Extension Jet. The absence of OME-A feedback inevitably leads to a reduction in eddy potential energy production in order to balance the energy budget, which results in a weakened mean current. The finding has important implications for improving climate models' representation of major oceanic fronts, which are essential components in the simulation and prediction of extratropical storms and other extreme events[3–6], as well as in the projection of the effect on these events of climate change.

With the availability of high-resolution satellite observations and climate model simulations, it has become increasingly evident that ocean mesoscale eddies and fronts can have a profound influence on the atmospheric planetary boundary layer[7–11]. In critical regions where prominent oceanic fronts reside, such as the Kuroshio and Gulf Stream Extensions, studies have demonstrated that the strong coupling between the atmosphere and ocean at meso- and frontal scales can be vital in anchoring atmospheric storm tracks[3,4,12,13] and providing a source of influence on mid-latitude climate variability and weather pattern changes[5,6,14]. Therefore, it is imperative that we understand the fundamental dynamics for maintaining oceanic frontal systems and improve climate models' ability to simulate and predict these important oceanic features.

In classic ocean circulation theory, the Kuroshio Current and its extension, like other mid-latitude western boundary current regimes, are explained as a part of ocean circulation gyres maintained by the large-scale wind stress acting on the ocean's surface. Through the action of planetary waves, the energy input by the winds is transported westward and accumulated along the western boundary, giving rise to the intense western boundary current—the Kuroshio, which swings to the east from Japan to form the energetic Kuroshio Extension Jet (KEJ) along the boundary between the North Pacific subtropical and subpolar gyres. The KEJ is so energetic that it is hydrodynamically unstable, allowing vigorous ocean eddies to form in its vicinity by extracting energy from the mean current, which in turn can feed back onto the mean jet[15–17]. In this classic description, the effect of OME-A interaction in the Kuroshio Extension region is not taken into consideration. However, recent high-resolution satellite observations and coupled atmosphere–ocean model simulations indicate a large potential energy release from mesoscale oceanic fronts and eddies to the overlying atmosphere. Figure 1 shows the relationship between mesoscale sea surface temperatures (SSTs) and surface turbulent heat fluxes (THFs) over the Kuroshio Extension region, derived from high-resolution satellite observed SST and reanalysis surface THFs. Various high-resolution reanalysis products reveal a remarkably consistent SST–THF relationship, statistically significant with P value less than 0.01, which gives an approximately 40–56 W m^{-2} THF increase (into the atmosphere) per 1 °C SST increase. This relationship is well reproduced by an eddy-resolving Community Earth System Model (CESM) simulation developed and conducted by the National Center for Atmospheric Research (see Methods). It is also well reproduced by an ensemble of higher-resolution Coupled Regional Climate Model (CRCM) simulations developed and performed at Texas A&M University (see Methods), except that the sensitivity of THF to SST change is at a higher rate of 56 W m^{-2} °C^{-1}. Whether this strong OME-A feedback will have an impact on the KEJ and the associated oceanic front is the focus of this study.

To examine this question, we conduct a set of twin simulations using both global and regional climate models (CESM and CRCM). The twin simulations consist of a standard high-resolution model simulation (the control simulation or CTRL) and an otherwise identical simulation except that the influence of mesoscale SST variability on the atmosphere is suppressed (hereafter referred to as the Mesoscale Eddy Filtered Simulation or MEFS). The suppression of mesoscale SST influence in MEFS is achieved by spatially low-pass-filtering the simulated SST that is given to the atmospheric model at each coupling step, so that the atmosphere does not 'feel' ocean eddies even though they are simulated by the ocean model (see Methods). As such, OME-A feedback is greatly reduced in MEFS and a comparison between CTRL and MEFS allows us to assess the impact of OME-A feedback on the KEJ. The conceptual design of the experiments is similar to a previous study of the effects of the eddies on the Gulf Stream[18], except that here a fully coupled—instead of an ocean-only—modelling framework is employed to allow more explicit and accurate simulations of the OME-A feedback.

Figure 2a and b displays the vertically averaged mean kinetic energy of the upper 250 m of the ocean from the 11-year global CESM CTRL

[1]Physical Oceanography Laboratory/CIMST, Ocean University of China and Qingdao National Laboratory for Marine Science and Technology, Qingdao, China. [2]Department of Oceanography, Texas A&M University, College Station, Texas, USA. [3]Department of Atmospheric Sciences, Texas A&M University, College Station, Texas, USA. [4]Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado, USA. [5]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany. [6]Faculty of Mathematics and Natural Sciences, University of Kiel, Kiel, Germany.

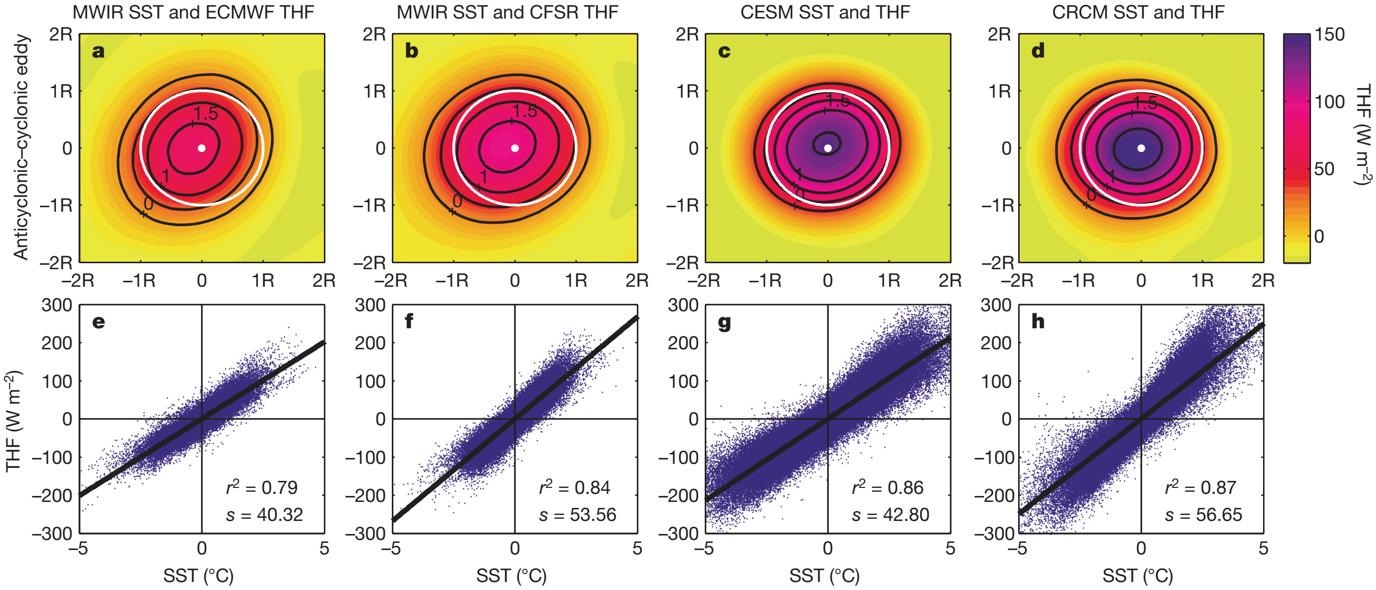


Figure 1 | Mesoscale surface THF and SST relationship. a–d, Composites of mesoscale-eddy-induced SST (contour) and THF (colour scale). Shown are winter season (ONDJFM) mean anticyclonic warm eddy minus cyclonic cold eddy composites in the Kuroshio Extension region (30° N–50° N, 140° E–180° E). The white circle marks one eddy radius (R) and the white point indicates the eddy centre. Daily surface heat fluxes from the European Centre for Medium-Range Weather Forecasts Year of Tropical Convection (ECMWF-YOTC, 2008–2010) and from the National Centers for Environmental Prediction–Climate Forecast System Reanalysis (NCEP–CFSR, 2002–2010) are used in **a** and **b**, respectively, together with satellite microwave-infrared (MWIR) SST over the overlapping periods. Simulated winter daily mean SST and THF from CESM CTRL and CRCM CTRL are used in **c** and **d**, respectively. The same eddy tracking algorithm was applied to the model simulated and observed sea surface height. See Methods for details of eddy tracking, composite and data sets used. **e**–**h**, Scatter plot and linear regression (black lines) between high-pass filtered weekly SST and THF in the same region as for the ECMWF-YOTC (**e**), NCEP-CFSR (**f**), CESM CTRL (**g**) and CRCM CTRL (**h**). r^2 represents the explained THF variance and *s* is the regression coefficient. The *P* values for all the regression coefficients are below 0.01, on the basis of an *F*-test.

and MEFS runs over the North Pacific. It is evident that the suppression of OME-A feedback leads to a more meandering and weaker KEJ (Methods and Extended Data Fig. 1). The difference between the mean kinetic energy in CTRL and MEFS clearly shows that over the entire North Pacific Ocean it is the KEJ that is most affected by the change in OME-A feedback and the more meandering KEJ in MEFS appears to shift northward (Fig. 2c). The more meandering and northward-shifted KEJ is consistent with the observations that the KEJ axis tends to move

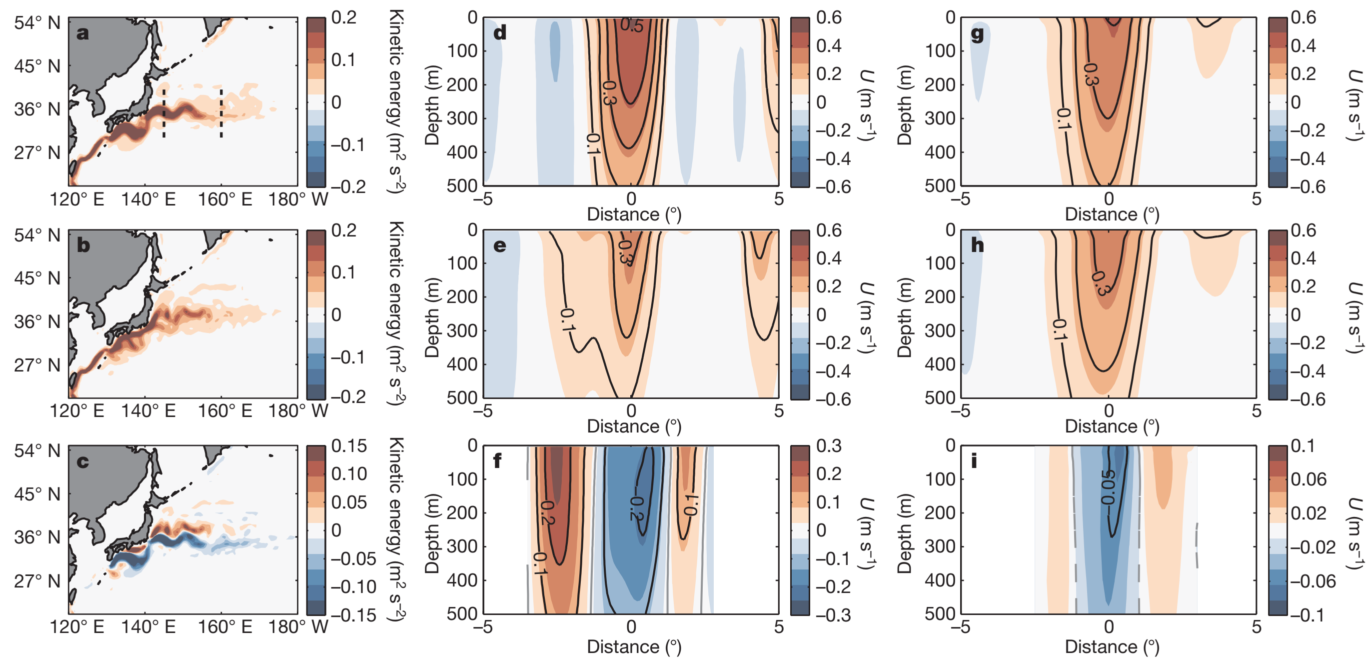


Figure 2 | Kuroshio response to OME-A feedback simulated by eddy-resolving global and regional climate models. a–c, Vertically averaged (upper 250 m) winter-season mean kinetic energy (colour scale) simulated in global CESM CTRL (**a**), MEFS (**b**) and MEFS minus CTRL (**c**) based on the winter-season (ONDJFM) mean of 11-year CESM simulations. **d–f**, Vertical section of zonally averaged (145° E–160° E) (indicated by the dashed vertical lines in **a**) winter-season mean KEJ velocity *U* (colour scale and contours in metres per second) across the Kuroshio Extension Front based on first-year CESM simulations. The *x* axis is the distance from the axis of KEJ at 0, which is defined by the latitude of the winter-season mean maximum *U*, and the *y* axis is depth. **g–i**, As for **d–f**, but for the ensemble mean of zonally averaged winter-season mean *U* velocity in regional CRCM simulations.

northward when the KEJ resides in an unstable (more meandering) regime[19]. Furthermore, the KEJ change occurs within a remarkably short time period, which is clearly illustrated by the sharp contrast between the zonal mean KEJ averaged over the first year of the CESM CTRL and MEFS runs: the KEJ in MEFS is considerably weaker and broader than that in CTRL (Fig. 2d–f), similar to the long-term averaged KEJ change (Extended Data Fig. 1). The rapid KEJ response to OME-A feedback changes is corroborated by the ensemble of regional CRCM runs that show a very similar weakening in the KEJ after OME-A feedback is suppressed (Fig. 2g–i), indicating the robustness of the finding. Because oceanic adjustment timescales to large-scale wind forcing through planetary waves are on the order of several years to decades at mid-latitudes[19], it is unlikely that this rapid KEJ response is caused by large-scale atmospheric circulation changes (Methods and Extended Data Fig. 2).

A further probe into the model response to OME-A feedback reveals that the weakened KEJ is accompanied with a substantial increase in mesoscale eddy energy when OME-A feedback is suppressed in MEFS. Figure 3 shows the ratio (MEFS/CTRL) between wavenumber spectra of eddy kinetic energy (EKE) derived from the suppressed-feedback and full-feedback simulations in the Kuroshio Extension region, as well as the ratio between wavenumber spectra of enstrophy (half of the vorticity variance), for both global CESM and regional CRCM experiments. Both model experiments show an increase in EKE and enstrophy for wavelengths less than about 500 km when comparing MEFS to CTRL, but the largest increase occurs at wavelengths less than 100 km (Fig. 3). In particular, the global CESM simulations exhibit a nearly 30% EKE increase and a 50% enstrophy increase at this mesoscale eddy range, while the increases in the regional CRCM twin ensembles are slightly smaller, probably owing to the short integration period. The energized mesoscale eddy activity implies a change in eddy energy balance; we therefore asked whether the KEJ response to OME-A feedback change can be explained in terms of the change in mesoscale eddy energetics.

To shed light on this question, we perform an ocean eddy potential energy (EPE) analysis using the twin ensembles of regional CRCM simulations (see Methods and Extended Data Fig. 3 for details). The rate of EPE change is determined by the interplay between energy production and destruction processes (see equation (1) in Methods). The production is primarily achieved through converting the mean available potential energy (MAPE) stored in the mean flow to EPE (hereafter referred to as MAPE-to-EPE conversion). The destruction consists of two processes: one converts EPE to EKE to energize eddy circulation (hereafter referred to as EPE-to-EKE conversion) and the other acts to dissipate EPE through either oceanic mixing or exporting eddy energy to the atmosphere (hereafter referred to as EPE dissipation)[18]. Thus, it is the EPE dissipation that is strongly affected by OME-A feedback.

To quantify the extent to which OME-A feedback can affect the EPE dissipation, we analyse and compare the vertically integrated EPE budget over the upper 250 m of the ocean in the twin ensembles of regional CRCM simulations (inset tables in Extended Data Fig. 3). With full OME-A feedback, CRCM CTRL reveals that nearly 74% of the EPE extracted from MAPE is lost owing to EPE dissipation and less than 22% is converted to EKE (the remaining 4% is accounted for by the EPE change during the integration period). This result clearly indicates that in the presence of OME-A feedback, nearly three-quarters of the Kuroshio EPE production is dissipated and only a small portion (less than a quarter) is used to drive eddy circulation. In sharp contrast, when OME-A feedback is suppressed, CRCM MEFS shows that the roles of EPE dissipation and of EPE-to-EKE conversion in the EPE budget are reversed: more EPE is being converted to EKE (about 47%) than being dissipated (about 40%). The increase in EPE-to-EKE conversion is consistent with the enhanced mesoscale eddy activity in MEFS as shown in Fig. 3. Thus, it is very clear that OME-A feedback is essential for dissipating ocean EPE and maintaining the EPE budget.

The change in EPE dissipation between the twin ensembles of regional CRCM simulations provides the key as to why the simulated KEJ behaves differently with and without OME-A feedback. With the full feedback, the Kuroshio system experiences strong dissipation. To generate and maintain an active eddy field, the KEJ must be sufficiently energetic to allow strong conversion from MAPE to EPE in order to overcome the vigorous EPE dissipation owing to OME-A feedback. With the suppressed OME-A feedback, EPE dissipation is greatly reduced, leaving the Kuroshio system in a less dissipative regime. In such a regime, more EPE is being converted to EKE than is being dissipated to the atmosphere, leading to a more energetic eddy field (Fig. 3) and a more meandering and weaker KEJ (Fig. 2 and Extended Data Fig. 4). A schematic summary of the proposed mechanism is shown in Fig. 4, which is supported by a detailed EPE budget analysis (Methods and Extended Data Fig. 3).

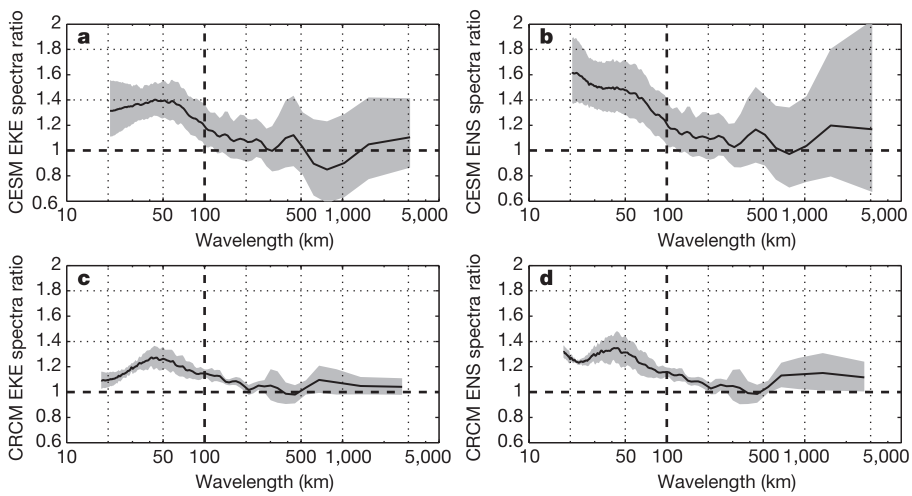


Figure 3 | Simulated energy and enstrophy spectra in the Kuroshio Extension region. a, **c**, Ratio (MEFS/CTRL) of winter-season (ONDJFM) mean EKE spectra in the Kuroshio Extension region (145° E–160° E, 30° N–42° N) based on the 11-year CESM simulations and the twin ensembles of CRCM simulations, respectively. **b**, **d**, As for **a** and **c**, but for eddy enstrophy (ENS; half of the vorticity variance) spectra. Eddy velocity is defined as the departure from the zonal mean between 145° E and 160° E. The shading indicates the geometric standard deviation of the EKE and the enstrophy ratio in individual years of the CESM simulations (**a** and **b**) and of CRCM simulations (**c** and **d**). The vertical dashed line represents eddies with 100-km wavelength. The horizontal dashed line indicates the MEFS/CTRL ratio of unity.

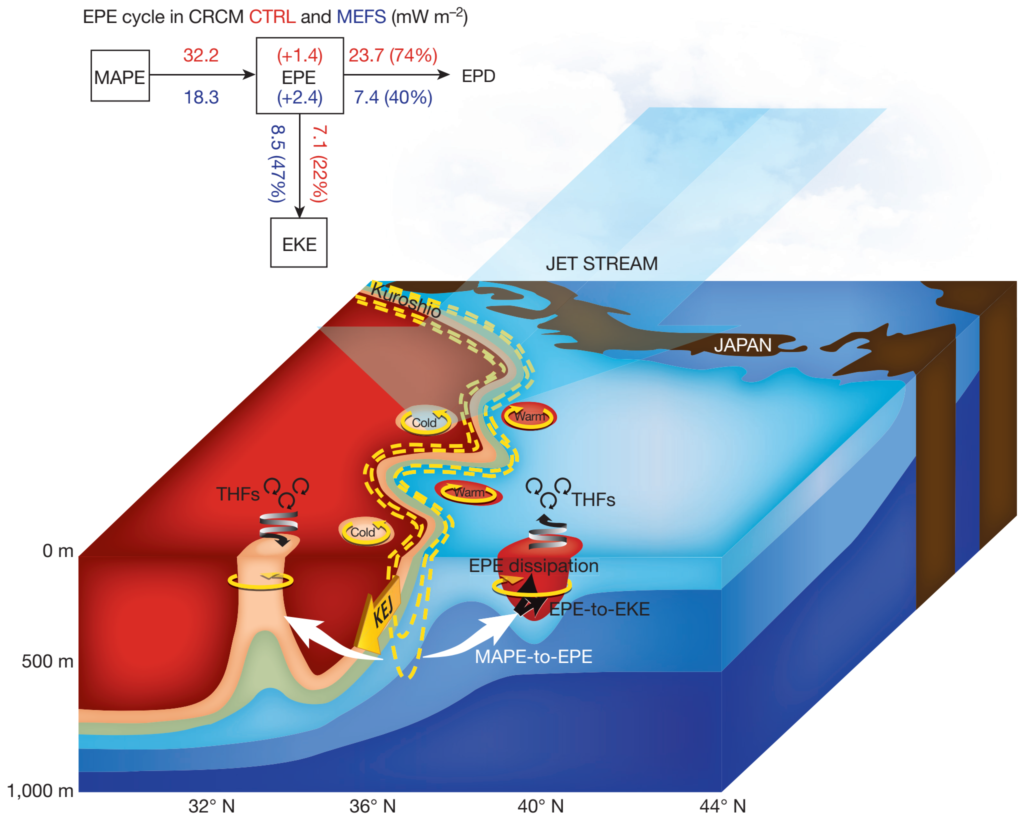


Figure 4 | Schematic of OME-A feedback regulating the KEJ. The Kuroshio Current swings to the east from Japan to form the energetic KEJ (yellow dashed contours), flowing eastward underneath the Pacific Jet Stream. The KEJ is hydrodynamically unstable, shedding eddies by converting KEJ's MAPE to EPE (white arrows). Energetic anticyclonic warm eddies (red circles) release THFs to the atmosphere (spiral arrows), and energetic cyclonic cold eddies (peach circles) absorb THFs from the atmosphere (spiral arrows), interacting with the overlying atmospheric planetary boundary layer and causing intense OME-A feedback. The feedback dominates EPE destruction (thicker black arrow) that dissipates 74% of the EPE generated from the KEJ in CRCM CTRL, leaving only a small portion (about 22%) to drive eddy circulation through EPE-to-EKE conversion (thinner black arrow). With suppressed feedback in CRCM MEFS, EPE dissipation (EPD) drastically reduces to 40% while EPE-to-EKE conversion increases to 47%, causing a more meandering and weaker KEJ. The EPE cycles in CTRL (red) and MEFS (blue) are illustrated in the inset diagram and further discussed in Methods and Extended Data Fig. 3.

These results from the high-resolution coupled atmosphere–ocean model experiments offer a new perspective on the dynamics controlling the Kuroshio Extension Front and Jet, revealing that local OME-A feedback needs to be taken into consideration by classic ocean circulation theory. This insight into the complex dynamics governing Kuroshio variability has important implications for the improvement of climate models in simulating the Kuroshio Extension Front and eddies as well as their impact on climate. Recent studies have shown that the sharp SST gradient and the eddy-induced mesoscale SST variability along the Kuroshio Extension Front can have an impact on the Pacific storm track[5,6,13], increasing the likelihood that an improved simulation of Kuroshio eddies and their feedback with the atmosphere in climate models may lead to improvements in short-term prediction of extratropical weather patterns and long-term projection of the storm track response to future climate change. Although the focus of this study is on the KEJ, the findings of this study are applicable to other oceanic fronts, such as the Gulf Stream front (Methods and Extended Data Fig. 5) and probably also Southern Ocean fronts, where strong OME-A feedback exists[1,20–22]. Not only are these oceanic frontal zones hotspots of the oceans where extratropical ocean influence on the atmosphere may be particularly strong[23], they are also the regions where present climate models have great difficulties in reproducing observations[24–29]. The results presented here may shed light on how to understand and reduce model biases in these critical regions and improve climate simulations and projection[30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 April; accepted 2 June 2016.

1. Kwon, Y.-O. *et al.* Role of the Gulf Stream and Kuroshio–Oyashio systems in large-scale atmosphere–ocean interaction: a review. *J. Clim.* **23,** 3249–3281 (2010).
2. Small, R. J. *et al.* A new synoptic scale resolving global climate simulation using the Community Earth System Model. *J. Adv. Model. Earth Syst.* **6,** 1065–1094 (2014).
3. Nakamura, H., Sampe, T., Tanimoto, Y. & Shimpo, A. Observed associations among storm tracks, jet streams and midlatitude oceanic fronts. *Geophys. Monogr. Ser.* **147,** 329–345 (2004).
4. Joyce, T. M., Kwon, Y.-O. & Yu, L. On the relationship between synoptic wintertime atmospheric variability and path shifts in the Gulf Stream and the Kuroshio Extension. *J. Clim.* **22,** 3177–3192 (2009).
5. Frankignoul, C., Sennechael, N., Kwon, Y. O. & Alexander, M. A. Influence of the meridional shifts of the Kuroshio and the Oyashio Extensions on the atmospheric circulation. *J. Clim.* **24,** 762–777 (2011).
6. Ma, X. *et al.* Distant influence of Kuroshio eddies on North Pacific weather patterns? *Sci. Rep.* **5,** 17785 (2015).
7. Chelton, D. B., Schlax, M. G., Freilich, M. H. & Milliff, R. F. Satellite measurements reveal persistent small-scale features in ocean winds. *Science* **303,** 978–983 (2004).
8. Xie, S.-P. Satellite observations of cool ocean–atmosphere interaction. *Bull. Am. Meteorol. Soc.* **85,** 195–208 (2004).
9. Bryan, F. O. *et al.* Frontal scale air–sea interaction in high-resolution coupled climate models. *J. Clim.* **23,** 6277–6291 (2010).
10. Small, R. J. *et al.* Air–sea interaction over ocean fronts and eddies. *Dyn. Atmos. Oceans* **45,** 274–319 (2008).

11. Putrasahan, D. A., Miller, A. J. & Seo, H. Isolating mesoscale coupled ocean–atmosphere interactions in the Kuroshio Extension region. *Dyn. Atmos. Oceans* **63,** 60–78 (2013).
12. Small, R. J., Tomas, R. A. & Bryan, F. O. Storm track response to ocean fronts in a global high-resolution climate model. *Clim. Dyn.* **43,** 805–828 (2014).
13. O'Reilly, C. H. & Czaja, A. The response of the Pacific storm track and atmospheric circulation to Kuroshio Extension variability. *Q. J. R. Meteorol. Soc.* **141,** 52–66 (2015).
14. Zhou, G. D., Latif, M., Greatbatch, R. J. & Park, W. Atmospheric response to the North Pacific enabled by daily sea surface temperature variability. *Geophys. Res. Lett.* **42,** 7732–7739 (2015).
15. Qiu, B. & Chen, S. M. Eddy-mean flow interaction in the decadally modulating Kuroshio Extension system. *Deep Sea Res. Part II* **57,** 1098–1110 (2010).
16. Waterman, S., Hogg, N. G. & Jayne, S. R. Eddy-mean flow interaction in the Kuroshio Extension region. *J. Phys. Oceanogr.* **41,** 1182–1208 (2011).
17. Zhang, Z. G., Wang, W. & Qiu, B. Oceanic mass transport by mesoscale eddies. *Science* **345,** 322–324 (2014).
18. Zhai, X. M. & Greatbatch, R. J. Surface eddy diffusivity for heat in a model of the northwest Atlantic Ocean. *Geophys. Res. Lett.* **33,** L24611 (2006).
19. Qiu, B. & Chen, S. Variability of the Kuroshio Extension jet, recirculation gyre, and mesoscale eddies on decadal time scales. *J. Phys. Oceanogr.* **35,** 2090–2103 (2005).
20. Minobe, S., Kuwano-Yoshida, A., Komori, N., Xie, S. P. & Small, R. J. Influence of the Gulf Stream on the troposphere. *Nature* **452,** 206–209 (2008).
21. Kelly, K. A. *et al.* Western boundary currents and frontal air–sea interaction: Gulf Stream and Kuroshio Extension. *J. Clim.* **23,** 5644–5667 (2010).
22. Frenger, I., Gruber, N., Knutti, R. & Munnich, M. Imprint of Southern Ocean eddies on winds, clouds and rainfall. *Nat. Geosci.* **6,** 608–612 (2013).
23. Wu, L. X. *et al.* Enhanced warming over the global subtropical western boundary currents. *Nat. Clim. Chang.* **2,** 161–166 (2012).
24. Gent, P. R., Yeager, S. G., Neale, R. B., Levis, S. & Bailey, D. A. Improvements in a half degree atmosphere/land version of the CCSM. *Clim. Dyn.* **34,** 819–833 (2010).
25. McClean, J. L. *et al.* A prototype two-decade fully-coupled fine-resolution CCSM simulation. *Ocean Model.* **39,** 10–30 (2011).
26. Delworth, T. L. *et al.* Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Clim.* **25,** 2755–2781 (2012).
27. Kirtman, B. P. *et al.* Impact of ocean model resolution on CCSM climate simulations. *Clim. Dyn.* **39,** 1303–1328 (2012).
28. Wang, C. Z., Zhang, L. P., Lee, S. K., Wu, L. X. & Mechoso, C. R. A global perspective on CMIP5 climate model biases. *Nat. Clim. Chang.* **4,** 201–205 (2014).
29. Drews, A., Greatbatch, R. J., Ding, H., Latif, M. & Park, W. The use of a flow field correction technique for alleviating the North Atlantic cold bias with application to the Kiel Climate Model. *Ocean Dyn.* **65,** 1079–1093 (2015).
30. Griffies, S. M. *et al.* Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *J. Clim.* **28,** 952–977 (2015).

Acknowledgements This research is a collaboration between TAMU and OUC led by P.C. at TAMU, and is supported by US National Science Foundation grants AGS-1462127 and AGS-1067937, and National Oceanic and Atmospheric Administration grant NA11OAR4310154, as well as by China's National Basic Research Priorities Programme (2013CB956204 and 2014CB745000) and the Natural Science Foundation of China (41490644 and U1406401). P.C. was partially supported by the Excellence Cluster, Future Ocean, Kiel and the SFB754 and by the 2015 Francis Bretherton Visitorship during his sabbatical leave at GEOMAR and NCAR, respectively. The Texas Advanced Computing Center at The University of Texas at Austin and the Texas A&M High Performance Research Computing provided the computing resources we needed to perform our CRCM simulations.

Author Contributions X.M. conducted the CRCM simulations and performed most of the analyses. Z.J. carried out the eddy energetic analysis. P.C. (lead investigator) conceived the project and wrote the manuscript. X.L. assisted with data analyses. R.M. developed CRCM and its parallel Loess filter and provided assistance during model setup and use. R.J.S. and F.O.B. performed the global CESM simulations. R.J.G. and P.B. were involved in interpreting the results. D.W, X.L. and L.W. participated in the design of the study. All authors contributed to improving the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C. (ping@tamu.edu).

Reviewer Information *Nature* thanks K. Kelly and M. Petersen for their contribution to the peer review of this work.

METHODS

Model configuration and experiment design. The high-resolution eddy-resolving global climate simulations in this study are conducted by the National Center for Atmospheric Research using the Community Earth System Model (CESM). The model configuration includes the Community Atmosphere Model version 5 (CAM5) with a spectral element dynamical core, the Parallel Ocean Program version 2 (POP2), as well as sea ice and land models[2]. CAM5 and POP2 have horizontal resolutions of about 0.25° and about 0.1°, respectively. The ocean model provides SST and surface current velocity to the atmosphere model and receives updated fluxes from the atmosphere model calculated using the Large and Yeager[31] surface layer scheme every 6 h (refs 2 and 32). CESM CTRL was integrated for 100 years with 14 years of spin-up. Starting on 1 March of year 59, the twin MEFS was conducted with the same settings as CTRL except that a low-pass spatial filter was applied to the POP2 simulated SST before being passed to CAM5 at each coupling step to remove mesoscale SST variability. The ocean in turn was forced by the atmospheric surface fluxes, which contained no signature of the mesoscale SST. The low-pass filter used was a 1,000 km × 1,000 km rectangular boxcar filter. MEFS was integrated for 11 years. The twin simulation comparison shown in this study is based on the 11-year CESM CTRL and MEFS simulation from year 59 to year 70.

The Coupled Regional Climate Model (CRCM) was developed at Texas A&M University to perform simulations at high resolution. The model consists of a custom coupling software framework that allows frequent mass, momentum and energy exchanges at the interface between its ocean and atmosphere regional model components. CRCM can efficiently perform remapping of interface fields as well as spatial filtering of ocean surface quantities at runtime, and provides a set of different surface flux formulations. It includes the Weather Research and Forecasting[33] (WRF) Model developed by the National Center for Atmospheric Research as the atmospheric component and the Regional Ocean Modelling Systems (ROMS) developed by Rutgers University and UCLA as the oceanic component. The CRCM is configured with a common Arakawa-C grid shared by ROMS and WRF at 9-km horizontal resolution extending from 3.6° N to 66° N and from 99° E to 270° E to cover the entire North Pacific Ocean. WRF is divided into 30 vertical levels and its detailed configuration follows a recent study[6]. ROMS has 50 levels in a vertical terrain-following coordinate and includes a K-profile parameterization turbulent mixing closure scheme[34] for vertical mixing and a bi-harmonic horizontal Smagorinsky-like mixing for the momentum[35]. In CRCM simulations, the coupling strategy was similar to that of CESM except that WRF's flux formulations were used and WRF and ROMS were coupled hourly. The CRCM CTRL comprised an ensemble of five 6-month integrations, initialized on 1 October 2003, 2004, 2005, 2006 and 2007, respectively. Each of these initial conditions for ROMS was derived from a 6-year spin-up run, carried out using version 2 of the Coordinated Ocean-ice Reference Experiments (CORE-II)[31] as surface forcing and 5-day averaged Simple Ocean Data Assimilation (SODA)[36] output as lateral boundary conditions. WRF initial conditions were derived from the NCEP-Department of Energy AMIP-II Reanalysis (NCEP-II)[37]. The lateral boundary conditions for the CRCM CTRL ensemble were derived from 6-hourly NCEP-II reanalysis for WRF and 5-day average SODA output for ROMS. The set-up of the MEFS ensemble was identical to CTRL except that the ROMS-simulated SST was processed by an embedded low-pass Loess filter with a 15° (longitude) × 5° (latitude) cut-off wavelength[38] before being provided to WRF at each coupling step.

Weakened KEJ in CESM MEFS. The weakened KEJ in global CESM simulations when OME-A feedback is suppressed can be shown more clearly by a composite analysis of zonally averaged winter-season (ONDJFM) mean U over the Kuroshio Extension region. The composite was taken according to the KEJ axis defined by the latitude of maximum U in each winter. Extended Data Fig. 1 shows the 11-winter-season composite of zonal-mean U from CTRL, MEFS and MEFS minus CTRL. The long-term averaged KEJ change shown in Extended Data Fig. 1a–c is remarkably similar to the first-year KEJ change shown in Fig. 2d–f and to the ensemble mean KEJ change simulated by regional CRCM shown in Fig. 2g–i.

Influence of the large-scale atmospheric circulation. To further confirm that the rapid KEJ response is unlikely to be caused by large-scale atmospheric circulation changes, Extended Data Fig. 2a shows the annual mean large-scale wind stress curl averaged over the 11-year global CESM CTRL simulation, which is based on spatially low-pass filtered wind stresses using a 10° (longitude) × 10° (latitude) boxcar filter. As a part of the subtropical gyre in the North Pacific, the Kuroshio is driven by the large-scale negative wind stress curl generated by the tropical easterly trade winds and mid-latitude westerlies. In the global CESM MEFS when OME-A feedback is suppressed, this large-scale wind stress curl does not exhibit much change (Extended Data Fig. 2b). Similar results are found in the regional CRCM MEFS and CTRL. These findings further suggest that the substantial weakening of KEJ shown in the main text cannot be attributed to changes in large-scale atmospheric circulations and is more probably influenced by the change in the local OME-A feedback.

Energy budget for EPE. In this study, we use the twin ensembles of short-term (6-month winter season) regional CRCM integrations, instead of the longer-term global CESM simulations, for the energetic analysis, mainly because of a programmatic reason: output of high-resolution (both in space and time) oceanic variables required for a detailed eddy energetic analysis is currently unavailable for the existing global simulations. Additionally, there are a number of scientific justifications for using the short-term regional CRCM simulations. First, the longer-term simulations allow planetary wave adjustment to affect the KEJ response to changes in large-scale atmospheric forcing, making it difficult to evaluate the role of local OME-A feedback in Kuroshio eddy energetics (although in our case we have demonstrated above that this is not a factor). Second, the global filtering adopted in the CESM MEFS can potentially affect tropical ocean circulations, which in turn can remotely influence Kuroshio eddy energetics, again complicating the interpretation of the OME-A feedback effect on Kuroshio eddy energetics. These problems are remedied by the regional modelling approach. However, the short integration period in this approach prevents an examination of equilibrium-state eddy energetics. Nevertheless, an analysis of transient-state eddy energetics can still be enlightening in understanding the role of OME-A feedback in maintaining oceanic eddies.

In analysing the eddy energy budget, EPE is defined as the potential energy associated with zonally perturbed flows between 145° E–160° E. Sensitivity tests suggest that using different regions (for example, 150° E–165° E) does not substantially affect the EPE budget analysis. The governing equation for EPE[39] is:

$$\frac{\partial}{\partial t}\left\langle \rho_0 \frac{b'^2}{2N_r^2} \right\rangle = -\left\langle \boldsymbol{u}_h \cdot \boldsymbol{\nabla}_h \left(\frac{b'^2}{2N_r^2} \right) \right\rangle - \left\langle \rho_0 \frac{v'b'^2}{N_r^2} \frac{\partial \bar{b}}{\partial y} \right\rangle - \left\langle \rho_0 w'b' \right\rangle + \left\langle \mathrm{EPD} \right\rangle \quad (1)$$

where $\langle \ldots \rangle$ denotes the horizontal area average, b is the buoyancy, $\boldsymbol{u}_h = (u, v)$ is the horizontal velocity, w is the vertical velocity, $\boldsymbol{\nabla}_h = (\partial/\partial x, \partial/\partial y)$, $\rho_0 = 1{,}025\,\mathrm{kg\,m^{-3}}$ is the water density, and N_r is the buoyancy frequency associated with a predefined background potential density profile ρ_r that is chosen to be the area average of the time-mean potential density (referenced to the sea surface) in the analysis domain (145° E–160° E, 30° N–42° N). Variables denoted by primes and bars correspond to the zonal perturbed and mean values, respectively.

The term on the left-hand side of equation (1) is the time rate of change of EPE, which is determined by the production and destruction terms on the right-hand side of equation (1). EPE production comes from the two processes: advection of EPE (the first term on the right-hand side of equation (1)) and eddy buoyancy fluxes acting on the mean meridional buoyancy gradient (the second term on the right-hand side of equation (1)), the combined effect of which is primarily to convert MAPE to EPE and which is therefore referred to as baroclinic conversion (that is, MAPE-to-EPE conversion in the main text). Destruction of EPE is described by the third and fourth terms on the right-hand side of equation (1): the vertical eddy buoyancy flux that converts EPE to EKE to energize eddy circulation (that is, EPE-to-EKE conversion) and EPE dissipation (EPD) through either oceanic mixing or exporting eddy energy to the atmosphere, as depicted in the main text. These terms are computed using daily mean diagnostic output from the CRCM simulations and the equation is exactly balanced.

The result of EPE analyses from the twin ensembles of regional simulations in the Kuroshio Extension region is summarized in Extended Data Fig. 3. In the CTRL ensemble with full OME-A feedback, it is evident that within the upper 50 m the main balance in the EPE budget is between the baroclinic conversion and EPE dissipation, both of which peak near the surface and decrease exponentially with depth. The EPE-to-EKE conversion is negligibly smaller near the surface, but becomes a dominant factor below 50 m. The tendency term, although non-zero due to the short integration, in general makes a minor contribution to the EPE budget in the Kuroshio Extension region. Upon a vertical integration from the surface to 250m (inset tables in Extended Data Fig. 3), the CRCM CTRL ensemble reveals a 32.2 $\mathrm{mW\,m^{-2}}$ baroclinic conversion, a 23.7 $\mathrm{mW\,m^{-2}}$ EPE dissipation, a 7.1 $\mathrm{mW\,m^{-2}}$ EPE-to-EKE conversion, and a 1.4 $\mathrm{mW\,m^{-2}}$ time rate of change, confirming that in the presence of OME-A feedback, most of the Kuroshio EPE, rather than being converted to EKE, is dissipated as illustrated in Fig. 4.

Since OME-A feedback is expected to strongly affect EPE dissipation because of the close mesoscale THF–SST relationship shown in Fig. 1, it is no surprise that the EPE budget is greatly altered in the CRCM MEFS ensemble where OME-A feedback is suppressed. From Extended Data Fig. 3, it is abundantly clear that the EPE dissipation is reduced by approximately threefold, particularly near the surface, in MEFS compared to CTRL, confirming the role of OME-A feedback in removing EPE to the atmosphere. To keep the EPE budget balanced, the baroclinic conversion (production term) is also greatly reduced, albeit not as much as the EPE dissipation reduction. Interestingly, EPE-to-EKE conversion is even enhanced in MEFS, indicating that more EPE is converted to EKE in the absence of OME-A feedback. These changes are more conveniently summarized in terms of a vertically

integrated (from surface to 250 m depth) EPE budget (inset tables in Extended Data Fig. 3). Compared to CTRL, the most important change in MEFS occurs in EPE dissipation, which is reduced by approximately 69% to 7.4 mW m^{-2} (from 23.7 mW m^{-2}). EPE-to-EKE conversion increases by approximately 20% to 8.5 mW m^{-2} (from 7.1 mW m^{-2}); combining the EPE dissipation and the EPE-to-EKE conversion in MEFS gives rise to a total reduction of EPE destruction by approximately 48% to 15.9 mW m^{-2} (from 30.8 mW m^{-2}). Baroclinic conversion reduces by approximately 43% to 18.3 mW m^{-2} (from 32.2 mW m^{-2}). Finally, the rate of EPE change, though increased from 1.4 mW m^{-2} to 2.4 mW m^{-2}, is still minor compared to other terms. Therefore, in the absence of OME-A feedback, even though EPE production is reduced, there is actually more EPE being converted to EKE and less EPE being dissipated, supporting the proposed mechanism, illustrated in Fig. 4. The major reduction in EPE dissipation also causes an increase in EPE during the integration period, resulting in an enhanced rate of EPE change, although its contribution to the total EPE budget remains small.

Since EPE production by baroclinic conversion is through down-gradient eddy buoyancy fluxes acting on the mean buoyancy gradient[40,41], its major reduction in the absence of OME-A feedback implies that there is a decrease in either the eddy buoyancy fluxes or mean buoyancy gradient. The former is an unlikely candidate because mesoscale eddy activity is energized, that is, both EPE and EKE increase, in MEFS compared to CTRL. It means that a weakened mean buoyancy gradient is a more likely consequence of OME-A feedback suppression in MEFS. Further analyses indicate that salinity makes a secondary contribution to the change of mean buoyancy gradient, and thus the decrease in baroclinic conversion is mainly due to the weakened potential temperature gradient. Extended Data Fig. 4 shows a comparison of the meridional potential temperature gradient composite along the Kuroshio Extension in global and regional simulations. The composite was referenced to the KEJ axis in each winter, as in the composite analysis of zonal-mean U. As expected, the strongest temperature gradient collocates with the KEJ owing to geostrophy. When OME-A feedback is suppressed, the temperature front is weakened by about 15% in MEFS compared to CTRL, corresponding to the broader and weaker KEJ shown in Fig. 2. The weakened temperature front and KEJ along with the reduced baroclinic conversion support the argument that less EPE needs to be extracted from MAPE to sustain the more energetic eddy field when OME-A feedback is suppressed, owing to the much reduced EPE dissipation and enhanced EPE-to-EKE conversion.

Change of mean flow and eddy activity in the Gulf Stream. To examine whether similar changes occur in the Gulf Stream extension region, we also analysed change of mean flow and eddy activity in the Gulf Stream extension region in the global CESM CTRL and MEFS simulations. Similar to the KEJ response in the North Pacific, we found a substantial weakening (about 25%) of the Gulf Stream Extension current in global MEFS when OME-A feedback is suppressed (Extended Data Fig. 5a–c) and an increase of EKE and enstrophy (approximately 20%–40%) at mesoscale range (below about 500 km) in MEFS (Extended Data Fig. 5d, e). The consistent mean flow and eddy activity response to suppression of OME-A feedback in the Kuroshio Extension region and Gulf Stream Extension further supports our hypothesis that OME-A feedback may be essential in regulating western boundary current regimes.

Horizontal wavenumber spectra of EKE and enstrophy. The horizontal wavenumber spectra of EKE and enstrophy in Fig. 3 and Extended Data Fig. 5 are computed as

$$\Phi_{\mathrm{EKE}}(k_{\mathrm{H}})=k_{\mathrm{H}}\int_0^{2\pi}\frac{1}{2}|\tilde{u}(k_{\mathrm{H}}\cos\theta,k_{\mathrm{H}}\sin\theta)|^2+\frac{1}{2}|\tilde{v}(k_{\mathrm{H}}\cos\theta,k_{\mathrm{H}}\sin\theta)|^2\,\mathrm{d}\theta \quad (2)$$

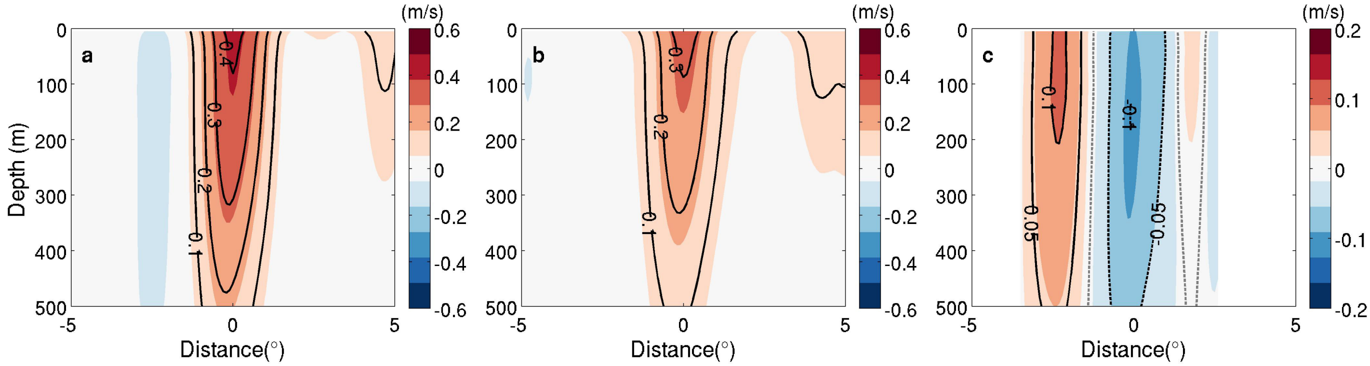
$$\Phi_{\mathrm{ENS}}(k_{\mathrm{H}})=k_{\mathrm{H}}\int_0^{2\pi}\frac{1}{2}|\tilde{\varsigma}(k_{\mathrm{H}}\cos\theta,k_{\mathrm{H}}\sin\theta)|^2\,\mathrm{d}\theta \quad (3)$$

where (u, v) are the zonal and meridional current components, ς is the vertical component of relative vorticity, k_{H} and θ are the magnitude and azimuth of the horizontal wave vector, and tilde represents the two-dimensional Fourier transform. Before computing the two-dimensional Fourier transform, a zonal mean is subtracted from daily mean (u, v) and ς to remove the background flow.

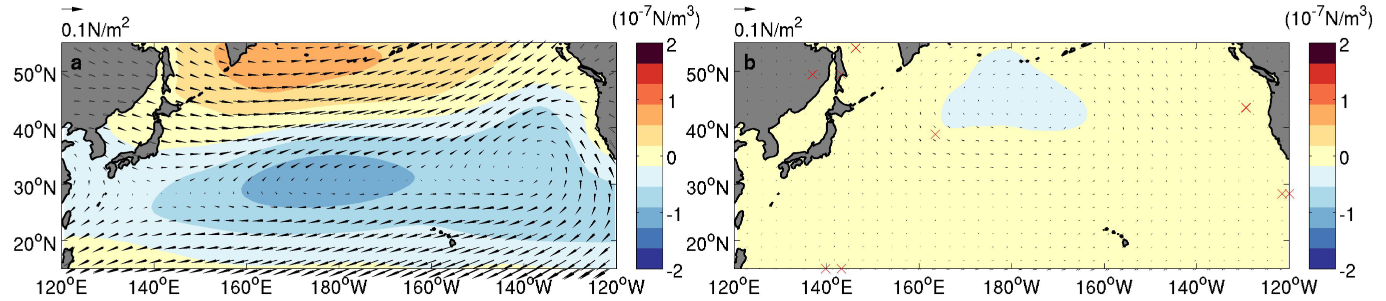
Data. The 0.25° daily satellite altimetry sea surface height data—AVISO[42] (http://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products.html) —are used to identify eddies in the Kuroshio Extension region combined with an eddy-tracking algorithm[43]. The European Centre for Medium-Range Weather Forecasts Year of Tropical Convection[44] (http://apps.ecmwf.int/datasets/) product at T799 spectral truncation (approximately 25 km) and the NCEP Climate Forecast System Reanalysis[45] (CFSR; http://cfs.ncep.noaa.gov/cfsr/downloads/) product at T385 spectral truncation (approximately 38 km) are used to derive the daily surface heat fluxes as shown in Fig. 1. The corresponding SST field is derived from 9-km daily satellite microwave-infrared SSTs (http://www.remss.com/measurements/sea-surface-temperature). Before making the eddy composite in Fig. 1, all eddies are normalized by their radius and rotated so that the background wind direction is directed to the east, following a recent study[22].

Code availability. The Community Earth System Model (CESM) developed by the National Center for Atmospheric Research can be downloaded online (http://www2.cesm.ucar.edu/). The Coupled Regional Climate Model (CRCM) developed at Texas A&M University can be obtained by written request to R.M. (rmontuoro@tamu.edu).

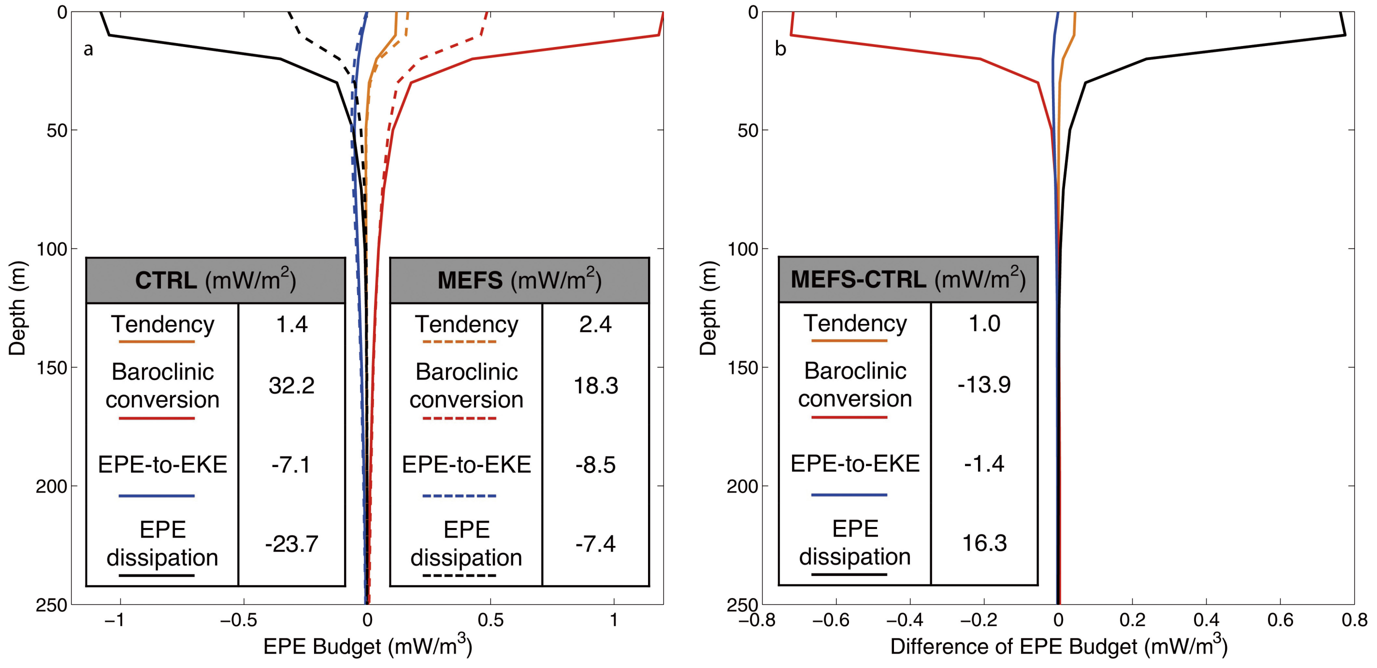
31. Large, W. G. & Yeager, S. G. The global climatology of an interannually varying air-sea flux data set. *Clim. Dyn.* **33,** 341–364 (2009).
32. Bishop, S. P., Bryan, F. O. & Small, R. J. Bjerknes-like compensation in the wintertime North Pacific. *J. Phys. Oceanogr.* **45,** 1339–1355 (2015).
33. Skamarock,W. C. *et al. A Description of the Advanced Research WRF Version 3.* NCAR Technical Note NCAR/TN-475+STR, http://dx.doi.org/10.5065/D68S4MVH (National Center for Atmospheric Research, 2008).
34. Large, W. G., Mcwilliams, J. C. & Doney, S. C. Oceanic vertical mixing—a review and a model with a nonlocal boundary-layer parameterization. *Rev. Geophys.* **32,** 363–403 (1994).
35. Griffies, S. M. & Hallberg, R. W. Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models. *Mon. Weath. Rev.* **128,** 2935–2946 (2000).
36. Carton, J. A. & Giese, B. S. A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA). *Mon. Weath. Rev.* **136,** 2999–3017 (2008).
37. Kanamitsu, M. *et al.* NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorol. Soc.* **83,** 1631–1643 (2002).
38. Schlax, M. G. & Chelton, D. B. Frequency-domain diagnostics for linear smoothers. *J. Am. Stat. Assoc.* **87,** 1070–1081 (1992).
39. von Storch, J. S. *et al.* An estimate of the Lorenz energy cycle for the world ocean based on the 1/10 degrees STORM/NCEP simulation. *J. Phys. Oceanogr.* **42,** 2185–2205 (2012).
40. Marshall, J. & Shutts, G. A note on rotational and divergent eddy fluxes. *J. Phys. Oceanogr.* **11,** 1677–1680 (1981).
41. Eden, C., Greatbatch, R. J. & Olbers, D. Interpreting eddy fluxes. *J. Phys. Oceanogr.* **37,** 1282–1296 (2007).
42. Ducet, N., Le Traon, P. Y. & Reverdin, G. Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. *J. Geophys. Res.* **105,** 19477 (2000).
43. Kurian, J., Colas, F., Capet, X., McWilliams, J. C. & Chelton, D. B. Eddy properties in the California Current System. *J. Geophys. Res.* **116,** C08027 (2011).
44. Waliser, D. E. *et al.* The "year" of tropical convection (May 2008-April 2010) climate variability and weather highlights. *Bull. Am. Meteorol. Soc.* **93,** 1189–1218 (2012).
45. Saha, S. *et al.* The NCEP climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.* **91,** 1015–1057 (2010).



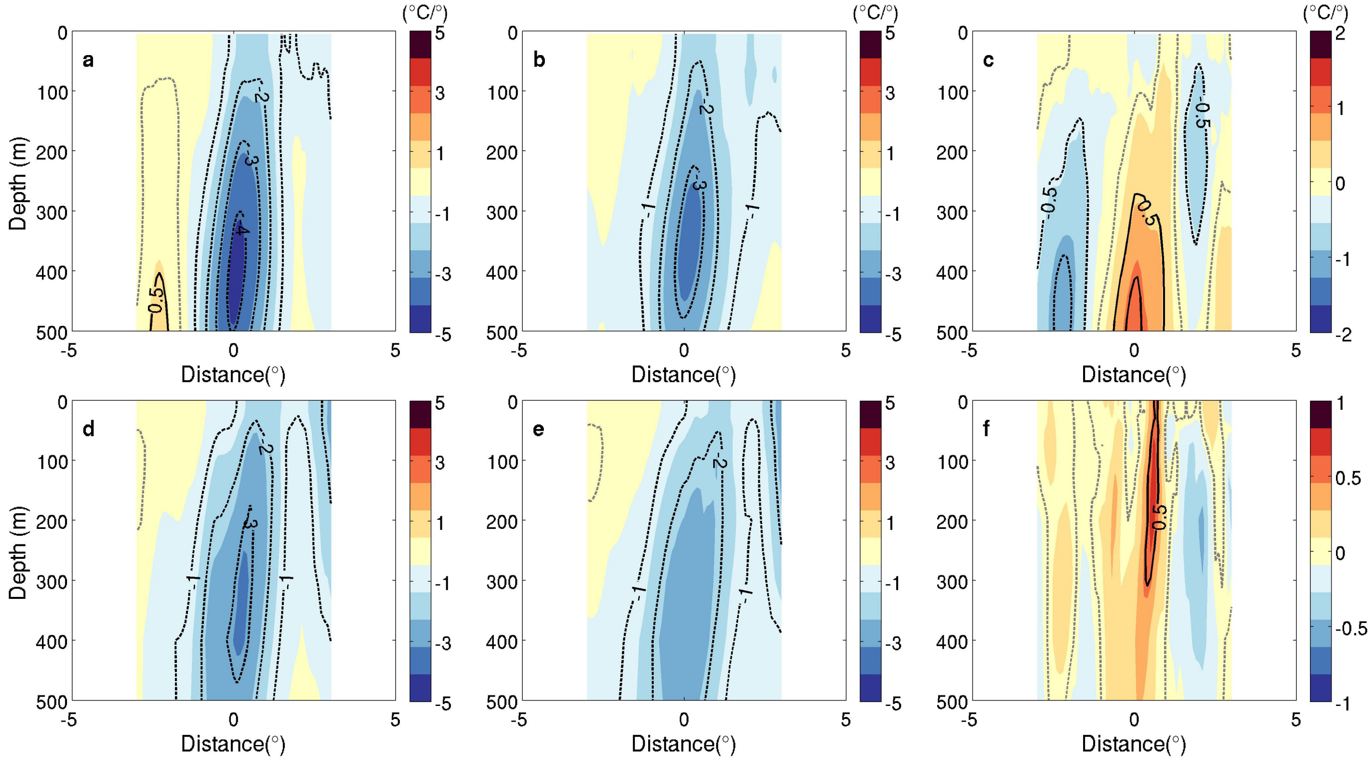
Extended Data Figure 1 | Long-term mean Kuroshio response to OME-A feedback simulated by CESM. Composite of winter-season (ONDJFM) mean *U* velocity (colour scale and contours in metres per second) zonally averaged between 145° E and 160° E for the whole of the 11-year CESM simulations in CTRL (**a**), MEFS (**b**) and MEFS minus CTRL (**c**). The composite was taken according to the axis of the winter mean KEJ defined by the latitude of maximum *U* velocity. The *x* axis is the distance from the KEJ axis.



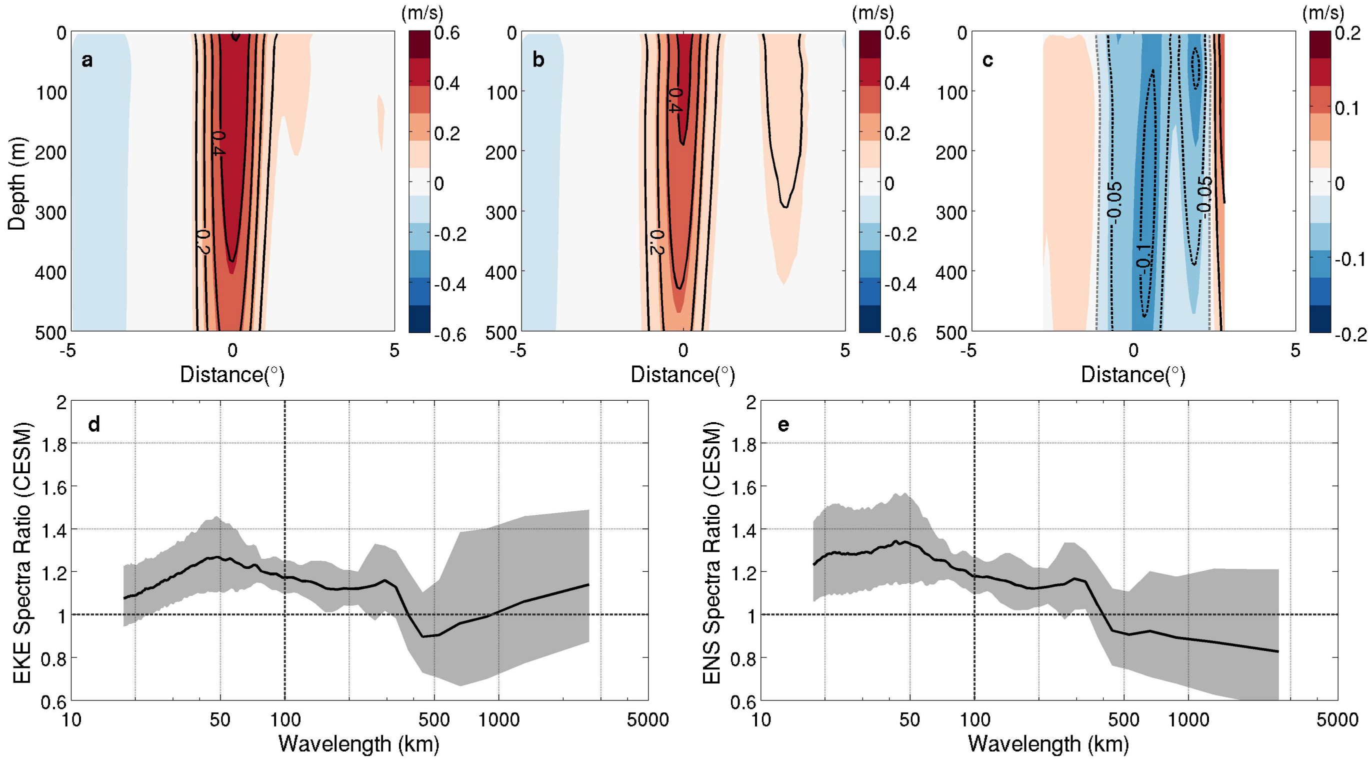
Extended Data Figure 2 | Large-scale wind stress and curl change simulated by CESM. Annual mean large-scale wind stresses (vectors) and wind stress curl (colour scale) in CTRL (**a**) and MEFS minus CTRL (**b**) averaged over the 11-year CESM simulations. The large-scale wind stresses are obtained using a 10° (longitude) × 10° (latitude) boxcar filter. The wind stress curl differences, significant with P value less than 0.05, based on monthly mean data using a Student's t-test are marked by red crosses in **b**.



Extended Data Figure 3 | EPE budget in the Kuroshio Extension region. **a**, Vertical profile of winter season (ONDJFM) mean EPE budget averaged over 145° E–160° E, 30° N–42° N for the CRCM CTRL ensemble (solid) and CRCM MEFS ensemble (dashed). **b**, As for **a**, except for MEFS minus CTRL. The inset tables summarize the vertically integrated EPE budget in the upper 250 m for CTRL and MEFS in **a** and for MEFS minus CTRL in **b**. Each of the EPE budget terms is indicated using a different colour, and solid and dashed lines in **a** represent CTRL and MEFS, respectively.



Extended Data Figure 4 | Temperature gradient response in the Kuroshio Extension region simulated by CESM and CRCM. **a**–**c**, Composite of winter-season mean meridional temperature gradient (colour scale and contours in degrees Celsius per degree latitude) averaged between 145° E and 160° E across the Kuroshio Extension Front in CTRL (**a**), MEFS (**b**) and MEFS minus CTRL (**c**) based on 11-year CESM simulations. The *x* axis is the distance from the KEJ axis at 0 defined by the latitude of winter mean maximum *U* and the *y* axis is depth. **d**–**f**, As for **a**–**c**, but for the CRCM twin ensembles.



Extended Data Figure 5 | Mean flow and eddy response to OME-A feedback in the Gulf Stream simulated by CESM. a–c, Composite of winter season mean *U* velocity (colour scale and contours in metres per second) along 60° W based on 11-year CESM simulations. The composite was taken according to the axis of the winter mean Gulf Stream current along 60° W defined by the latitude of maximum *U* velocity. The *x* axis is the distance from the Gulf Stream current. **d**, Ratio of EKE spectra between MEFS and CTRL in the Gulf Stream Extension region (35° N–42° N, 70° W–55° W) based on 11-year CESM simulations. **e**, As for **d**, but for eddy enstrophy (ENS) spectra. The grey shading indicates the geometric standard deviation of EKE and enstrophy ratio in individual years of the CESM simulations (**d** and **e**).

LETTER

doi:10.1038/nature18012

High-resolution seismic constraints on flow dynamics in the oceanic asthenosphere

Pei-Ying Patty Lin[1]†, James B. Gaherty[1], Ge Jin[1], John A. Collins[2], Daniel Lizarralde[2], Rob. L. Evans[2] & Greg Hirth[3]

Convective flow in the mantle and the motions of tectonic plates produce deformation of Earth's interior, and the rock fabric produced by this deformation can be discerned using the anisotropy of the seismic wavespeed[1–3]. This deformation is commonly inferred close to lithospheric boundaries beneath the ocean in the uppermost mantle, including near seafloor-spreading centres as new plates are formed via corner flow[4], and within a weak asthenosphere that lubricates large-scale plate-driven flow and accommodates smaller-scale convection[5,6]. Seismic models of oceanic upper mantle differ as to the relative importance of these deformation processes: seafloor-spreading fabric is very strong just beneath the crust–mantle boundary (the Mohorovičić discontinuity, or Moho) at relatively local scales[7,8], but at the global and ocean-basin scales, oceanic lithosphere typically appears weakly anisotropic when compared to the asthenosphere[9,10]. Here we use Rayleigh waves, recorded across an ocean-bottom seismograph array in the central Pacific Ocean (the NoMelt Experiment), to provide unique localized constraints on seismic anisotropy within the oceanic lithosphere–asthenosphere system in the middle of a plate. We find that azimuthal anisotropy is strongest within the high-seismic-velocity lid, with the fast direction coincident with seafloor spreading. A minimum in the magnitude of azimuthal anisotropy occurs within the middle of the seismic low-velocity zone, and then increases with depth below the weakest portion of the asthenosphere. At no depth does the fast direction correlate with the apparent plate motion. Our results suggest that the highest strain deformation in the shallow oceanic mantle occurs during corner flow at the ridge axis, and via pressure-driven or buoyancy-driven flow within the asthenosphere. Shear associated with motion of the plate over the underlying asthenosphere, if present, is weak compared to these other processes.

The NoMelt experiment was designed to provide localized constraints on the lithosphere–asthenosphere system in a normal oceanic tectonic plate, far from the influence of melting, either at the ridge or due to subsequent intraplate volcanism. NoMelt is centred on mature (70 million years old) Pacific sea floor southeast of Hawaii, between the Clarion and Clipperton fracture zones (Fig. 1). It has an aperture of 600 km × 400 km and consists of a broadband ocean-bottom seismograph (OBS) seismic array, a long-period magnetotelluric survey, and an active-source seismic refraction/reflection survey using a short-period OBS array. The year-long broadband OBS array recorded a large collection of shallow-source earthquakes from around the Pacific basin, and the vertical-component recordings of the Rayleigh surface waves from these events are exceptionally high quality within the period band 20–160 s, once the effects of ocean infragravity waves and tilt noise are removed (see Methods) (Extended Data Fig. 1). By applying ambient-noise interferometry, we extend the Rayleigh-wave observations into a shorter-period (10–20 s) band.

Surface waves traverse NoMelt at a wide range of back-azimuths, providing exceptional azimuthal coverage for estimating shear velocities and azimuthal anisotropy within the array footprint. We utilize a wide-band intra-array cross-correlation analysis to measure Rayleigh-wave phase velocities in the earthquake band (20–160 s), and a spectral fitting procedure to estimate phase velocity in the ambient-noise band (10–20 s) (see Methods). We find minimal lateral variation in phase velocity within the array, and the observed phase velocities can be expressed by a single, average isotropic phase velocity curve for the entire array, plus a periodic $\cos(2\theta)$ variation in phase velocity as a function of azimuthal direction θ for each frequency (Fig. 2). Such

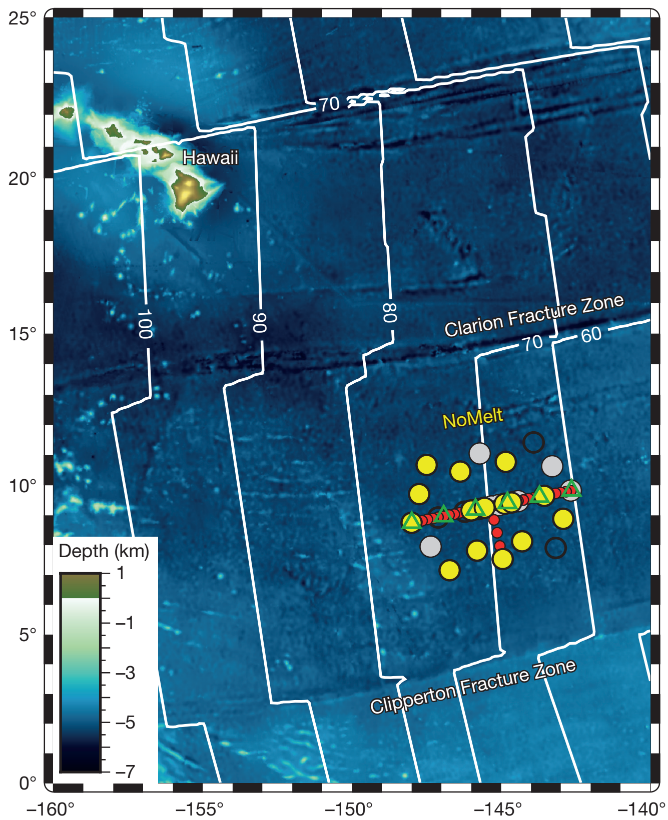


Figure 1 | Layout of NoMelt deployment on a bathymetric map of the central Pacific, with seafloor age (in millions of years) shown as white contours. Solid yellow circles represent broadband OBSs that produced good data; open and grey circles show locations of instruments that were unrecovered or produced poor data, respectively. Red circles and green triangles show locations of short-period OBSs and magnetotelluric instruments, for reference. Seafloor bathymetry was extracted from NOAA (https://www.ngdc.noaa.gov/mgg/global/global.html), and the seafloor age is available from EarthByte (http://www.earthbyte.org/Resources/Agegrid/1997/digit_isochrons.html).

[1]Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA. [2]Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA. [3]Geological Sciences Department, Brown University, Providence, Rhode Island, USA. †Present address: Taiwan Ocean Research Institute, National Applied Research Laboratories, Kaohsiung, Taiwan (P.-Y.P.L.).

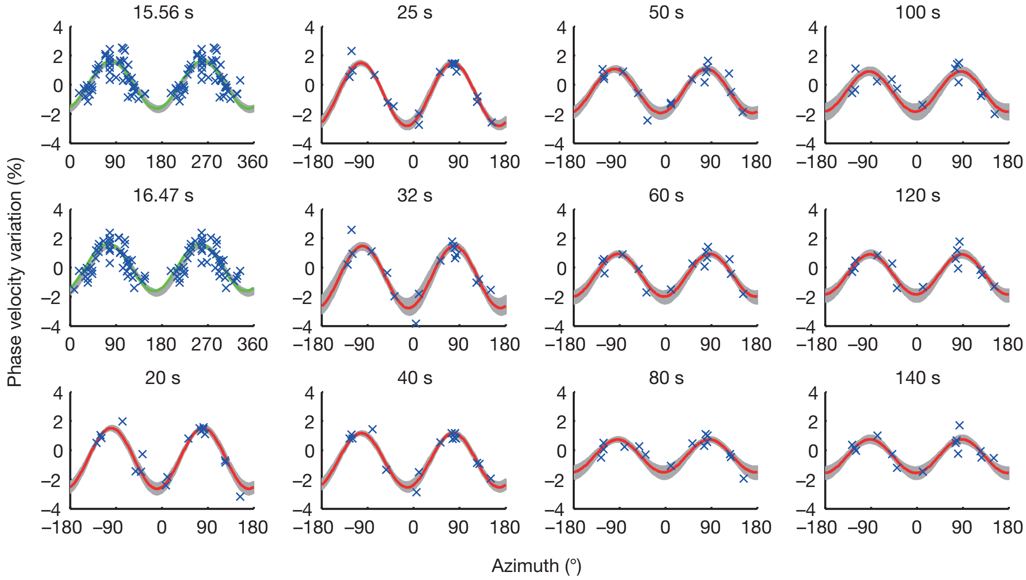


Figure 2 | Examples of the 2θ azimuthal variation of phase velocity measurements at different periods. Blue crosses indicate phase velocity measurements; green and red solid lines are the best-fitting 2θ variation for Rayleigh waves observed from ambient noise and teleseismic events, respectively, with grey bands representing the spread of acceptable 2θ fits as determined by a boot-strap resampling procedure. Period is shown for each panel.

azimuthal anisotropy is diagnostic of deformation-induced olivine fabric in the mantle[11], and we observe a very strong peak at ~80° azimuth at periods of 14–40 s. The anisotropy appears to weaken at intermediate periods, reaching a minimum at a period of about 80 s before strengthening and rotating slightly towards a more east–west direction at the longest periods (Fig. 2 and insets in Fig. 3b and c).

The average of all the phase velocity dispersion measurements can be inverted to estimate the isotropic one-dimensional vertically polarized shear-velocity structure (V_{sv}) beneath the array. We utilize an iterative, linearized least-squares inversion algorithm, with an additional Monte Carlo scheme to evaluate uncertainty associated with choice of starting model. The best-fitting V_{sv} profiles (Fig. 3a) are characterized by a thin high-velocity lid overlying a modest low-velocity zone, similar to other local and/or regional models of the Pacific[12,13]. The absolute velocities in both the lid and the low-velocity zone, as well as the transition between them (50–80 km depth), are very tightly constrained owing to the small uncertainty in the average phase velocity observations. The inferred seismic velocities are consistent with electrical conductivity profiles that suggest a dry lid overlying a modestly damp asthenosphere at subsolidus temperatures[14]—no melt appears to be required to produce either the low-velocity zone[15] or the high-conductivity zone[14], although very small amounts of melt cannot be ruled out. Regardless of melt content, the rheological profile across the system should correlate strongly with the shear-velocity profile, with a minimum in viscosity that coincides roughly with the middle of the low-velocity zone[16].

We invert the frequency-dependent cos(2θ) variations in phase velocity for 2θ models of azimuthal anisotropy (the G parameter) as a function of depth (Fig. 3b and c) using a smooth parameterization and a linearized least-squares algorithm (see Methods). The best-fitting models show the strongest anisotropy in the lithosphere with the fast direction parallel to the fossilized (dating to plate formation) seafloor-spreading direction (78°) estimated from the orientation of nearby fracture zones, and then decreasing in magnitude with depth into the asthenosphere. The weakest azimuthal anisotropy is centred in the asthenosphere at 100–150 km depth, and then it increases to a secondary maximum at about 250 km depth. Fast directions rotate modestly from the fossil-spreading direction within the lithosphere to a more east–west direction at depth. The data provide clear evidence for this basic layered structure (insets to Fig. 3b and c): strong G anisotropy at the Moho is driven by the peak in azimuthal velocity variations near 20-s period; the decay in G amplitude is required by the steep reduction in azimuthal variation in the period band 20–50 s; the secondary maximum in G amplitude near 250 km depth is required by the persistence of peak-to-peak azimuthal variations of ~3% out to the longest period; and the rotation in G azimuth with depth in the mantle is required by the rotation of the fast propagation direction from ~75° to ~90° across the period band 50–140 s. The preferred models are not consistent with anisotropy in the asthenosphere aligned with apparent plate motion (APM) in a hotspot reference frame (Fig. 3c).

The models of azimuthal anisotropy have several key implications for the nature of deformation and flow associated with oceanic tectonic plates. First, corner flow at the mid-ocean ridge represents the dominant fabric-forming process in the shallow oceanic mantle. This mechanism has long been recognized as important for explaining olivine fabric observed in ophiolites[17] and abyssal peridotites[18] and the presumably related seismic anisotropy observed in the oceanic lithosphere[1], particularly at the shallow depths constrained by active-source seismic refraction experiments[7,8,19]. Our results suggest that this fabric extends down at least 50 km into the cooling plate, providing new constraints on models of ridge dynamics[4]. The strength of this anisotropy is noticeably greater than that associated with deformation in the underlying asthenosphere. This high-resolution result is at odds with Pacific-wide studies of seismic anisotropy, which argue for a weakly anisotropic lithosphere underlain by a highly anisotropic asthenosphere[9,20]. This difference probably reflects the limited ability of large-scale studies to resolve structure imparted by seafloor spreading[21], which exhibits lateral heterogeneity over short length scales, as compared to larger-scale processes such as apparent plate motion or density-driven flow[22,23].

Below the plate, our results suggest that the strongest deformation is induced by dynamic flow within the asthenosphere, rather than passive shear strain associated with motion of the plate over the underlying mantle. This conclusion is supported by the absence of an APM signal and by the pattern of anisotropic strength with depth: the fast anisotropic direction does not align with APM within any depth interval, and the anisotropy is weakest near the base of the plate, where APM-induced strain should accumulate, and peaks near 250 km depth, deep in the asthenosphere, consistent with deformation caused by dynamic flow. Although our results strongly favour asthenospheric flow as the source of the observed anisotropic signal beneath the plate, we cannot directly constrain the details of this flow. This is because the apparent fast directions correspond to differential shear between plate flow, flow within the asthenosphere, and flow within the subasthenospheric mantle[24]—the former is known, but the latter two are not. This ambiguity exists for the case of shear flow with length scales larger than the 400-km array, and for deformation processes with length scales smaller than the NoMelt array, where the azimuthal signal will additionally include averages of local fabric variations.

The lack of an APM signal in our high-resolution study contrasts sharply with most large-scale studies of the Pacific mantle[9–11,20]. The

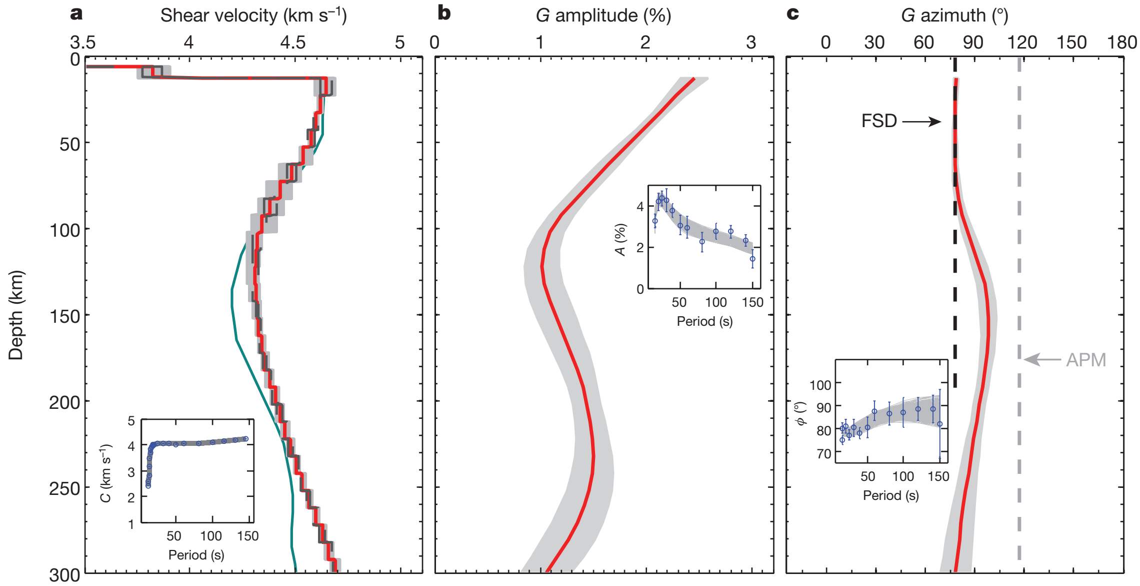


Figure 3 | Shear velocity and azimuthal anisotropy models in the upper mantle beneath NoMelt. a, The grey band depicts the span of shear velocity models resulting from inversion using 100 distinct starting models; the mean shear velocity model and associated standard deviation (s.d.) are shown by the solid red line and dashed black line, respectively. The blue line is a 52–110-million-year-old Pacific model[13] as a reference. The inset shows the phase velocities (*C*, calculated on the basis of the inverted shear velocity models; grey lines) compared to the observations (blue circles with 1 s.d.). **b**, The modelled strength of azimuthal anisotropy is shown by the solid red line with the grey shading showing ±1 s.d. from 1,000 randomly generated resamplings of the observed measurements. The inset shows the predicted peak-to-peak amplitude (*A*; grey lines) compared to the observations (blue circles with error bars of 1 s.d.). **c**, As for **b** but showing the direction of the azimuthal anisotropy. The vertical black and grey dashed lines indicate the orientation of fossil spread direction (FSD) and apparent plate motion (APM)[33]. The inset shows the predicted fast direction (ϕ; grey lines) compared to the observations (blue circles with error bars of 1 s.d.).

difference between our inferred fast directions and that predicted by APM is robust and statistically significant (see Methods). The discrepancy between our result and previous models probably arises due to the averaging inherent in seismic tomography; the amplitude of the largest-scale structures is recovered, whereas the strength of smaller-scale structures is underestimated if they are poorly resolved. APM is coherent at the scale of the entire Pacific plate, and this may explain why it is so prominent in Pacific-scale tomographic models. In detail, a few plate-scale models show subtle non-APM variations in the NoMelt region[3], and some authors prefer alternative flow processes to APM even at the plate scale[23]. Our result demonstrates that the smaller-scale processes of corner flow and asthenospheric dynamics are regionally more dominant than APM in controlling the development and evolution of seismic anisotropy in the mantle beneath the oceans.

The pattern (or structure) of anisotropy with depth suggests two scenarios for the dominant geodynamic flow in the central Pacific asthenosphere. Both scenarios hypothesize a low-viscosity asthenospheric channel[5,25], with the viscosity within the channel described by a power-law rheology. Large-scale pressure gradients within the asthenosphere would induce Poiseuille flow with a plug-like velocity profile (Fig. 4a). In such flow, fabric-inducing strain gradients are localized near the top and bottom boundaries of the channel, with relatively little fabric produced in the centre of the system; this scenario nicely matches the anisotropy–strength profile observed beneath NoMelt. Anisotropy

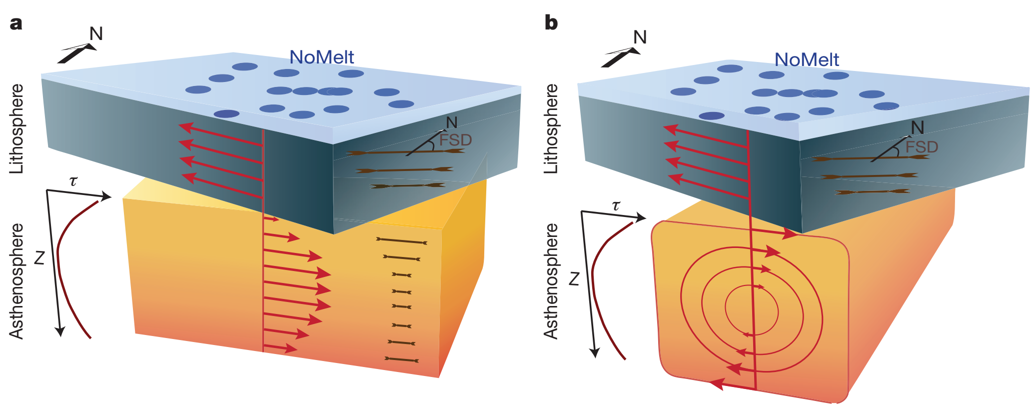


Figure 4 | Schematic of proposed models for azimuthal anisotropy beneath the NoMelt experiment. a, b, Pressure-gradient-driven flow (**a**) and density-driven small-scale convection (**b**) in the asthenosphere. At the top is the surface view; black arrows show the direction of north, with circles denoting NoMelt broadband OBSs. The red arrows indicate the flow direction in the lithosphere–asthenosphere system. The lithosphere moves at plate velocity at all depths. The direction of dynamic flow in the asthenosphere is shown as roughly east-to-west, as obtained from our azimuthal anisotropy model. The anisotropy signals in the lithosphere are frozen rock fabric (shown as brown segments) imparted by shear strain during corner flow at the ridge axis. Anisotropic strength in the lithosphere is largest at the top, decreases with depth, and is orientated with the FSD. The shear-strain gradient (maroon curve; Z, depth; τ, shear strain) resulting from dynamic flow is largest at the top and the base of the asthenosphere and smallest in the mid-asthenosphere for flow driven by either pressure gradients or density, producing rock fabric as indicated by the brown segments.

associated with the top boundary is difficult to distinguish from the lithospheric fabric, but the peak near 250 km depth would represent a shear zone at the lower boundary of the channel. Thermally induced pressure gradients within the asthenosphere may be associated with several geodynamic phenomena, including return flow from subduction zone to mid-ocean ridge[26], ridge-induced flow beneath the oceanic plates[27], and, more generally, density-driven flow[22,23]. Assuming a static upper mantle beneath the asthenosphere, the apparent east–west fast direction of anisotropy in the lower asthenosphere is generally consistent with the first two scenarios. That is, ridge-driven flow would be approximately east–west, while return flow would be west–east. Published density-driven flow models do not predict east–west asthenospheric flow in the NoMelt region[23] so the observations presented here should prove useful in refining future models.

Alternatively, density-driven small-scale convection within the low-viscosity channel could also produce fabric consistent with the anisotropy observations (Fig. 4b). In this scenario, horizontal flow at the top and bottom of convection cells produces strong azimuthal anisotropy for surface waves, while the centre of the channel exhibits weak and/or quasi-vertical fabric that results in little detectable azimuthal anisotropy, especially if the convection is on a length scale smaller than the array. Small-scale convection has long been invoked to explain bathymetry[28] and gravity observations[29] in the Pacific, most specifically in the form of organized thermal instabilities ("Richter Rolls"[30]) beneath the base of the plate. Direct observations of such structures in the asthenosphere have proved elusive[31], and our observation of east–west fabric is not consistent with the traditional Richter-Roll mechanism, which predicts horizontal fabric orthogonal to the APM-oriented long axis of the rolls. However, a recent high-resolution gravity compilation[32] reveals gravity lineations with orientations that vary over length scales of several hundred kilometres, perhaps indicating that small-scale convection beneath the Pacific is more three-dimensional than the traditionally invoked mechanism. Our observations may place direct constraints on the localized orientation of small-scale convection within the NoMelt region.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 October 2015; accepted 13 April 2016.
Published online 6 July 2016.

1. Hess, H. H. Seismic anisotropy of the uppermost mantle under oceans. *Nature* **203,** 629–631 (1964).
2. Forsyth, D. W. The early structural evolution and anisotropy of the oceanic upper mantle. *Geophys. J. Int.* **43,** 103–162 (1975).
3. Montagner, J. P. Upper mantle low anisotropy channels below the Pacific Plate. *Earth Planet. Sci. Lett.* **202,** 263–274 (2002).
4. Blackman, D. K. & Kendall, J. M. Seismic anisotropy in the upper mantle 2. Predictions for current plate boundary flow models. *Geochem. Geophys. Geosyst.* **3,** 8602 (2002).
5. Phipps Morgan, J., Morgan, W. J., Zhang, Y.-S. & Smith, W. H. F. Observational hints for a plume-fed, suboceanic asthenosphere and its role in mantle convection. *J. Geophys. Res.* **100,** 12753–12767 (1995).
6. Agrusta, R. *et al.* Small-scale convection in a plume-fed low-viscosity layer beneath a moving plate. *Geophys. J. Int.* **194,** 591–610 (2013).
7. Raitt, R. W., Shor, G. G., Francis, T. J. G. & Morris, G. B. Anisotropy of the Pacific upper mantle. *J. Geophys. Res. Solid Earth* **74,** 3095–3109 (1969).
8. Gaherty, J. B., Lizarralde, D., Collins, J. A., Hirth, G. & Kim, S. Mantle deformation during slow seafloor spreading constrained by observations of seismic anisotropy in the western Atlantic. *Earth Planet. Sci. Lett.* **228,** 255–265 (2004).
9. Beghein, C., Yuan, K., Schmerr, N. & Xing, Z. Changes in seismic anisotropy shed light on the nature of the Gutenberg discontinuity. *Science* **343,** 1237–1240 (2014).
10. Burgos, G. *et al.* Oceanic lithosphere-asthenosphere boundary from surface wave dispersion data. *J. Geophys. Res. Solid Earth* **119,** 1079–1093 (2014).
11. Montagner, J.-P. & Nataf, H.-C. Vectorial tomography—I. Theory. *Geophys. J. Int.* **94,** 295–307 (1988).
12. Tan, Y. & Helmberger, D. V. Trans-Pacific upper mantle shear velocity structure. *J. Geophys. Res.* **112,** B08301 (2007).
13. Nishimura, C. E. & Forsyth, D. W. The anisotropic structure of the upper mantle in the Pacific. *Geophys. J. Int.* **96,** 203–229 (1989).
14. Sarafian, E. *et al.* The electrical structure of the central Pacific upper mantle constrained by the NoMelt experiment. *Geochem. Geophys. Geosyst.* **16,** 1115–1132 (2015).
15. Jackson, I. & Faul, U. H. Grainsize-sensitive viscoelastic relaxation in olivine: towards a robust laboratory-based model for seismological application. *Phys. Earth Planet. Inter.* **183,** 151–163 (2010).
16. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144,** 93–108 (1996).
17. Nicolas, A., Ceuleneer, G., Boudier, F. & Misseri, M. Structural mapping in the Oman ophiolites: mantle diapirism along an oceanic ridge. *Tectonophysics* **151,** 27–56 (1988).
18. Ismaïl, W. B. & Mainprice, D. An olivine fabric database: an overview of upper mantle fabrics and seismic anisotropy. *Tectonophysics* **296,** 145–157 (1998).
19. Shearer, P. & Orcutt, J. Anisotropy in the oceanic lithosphere — theory and observations from the Ngendei seismic refraction experiment in the south-west Pacific. *Geophys. J. R. Astron. Soc.* **80,** 493–526 (1985).
20. Maggi, A., Debayle, E., Priestley, K. & Barruol, G. Azimuthal anisotropy of the Pacific region. *Earth Planet. Sci. Lett.* **250,** 53–71 (2006).
21. Debayle, E. & Ricard, Y. Seismic observations of large-scale deformation at the bottom of fast-moving plates. *Earth Planet. Sci. Lett.* **376,** 165–177 (2013).
22. Conrad, C. P. & Behn, M. D. Constraints on lithosphere net rotation and asthenospheric viscosity from global mantle flow models and seismic anisotropy. *Geochem. Geophys. Geosyst.* **11,** Q05W05 (2010).
23. Becker, T. W., Conrad, C. P., Schaeffer, A. J. & Lebedev, S. Origin of azimuthal seismic anisotropy in oceanic plates and mantle. *Earth Planet. Sci. Lett.* **401,** 236–250 (2014).
24. Silver, P. G. & Holt, W. E. The mantle flow field beneath western North America. *Science* **295,** 1054–1057 (2002).
25. Becker, T. W. On the effect of temperature and strain-rate dependent viscosity on global mantle flow, net rotation, and plate-driving forces. *Geophys. J. Int.* **167,** 943–957 (2006).
26. Buck, W. R., Small, C. & Ryan, W. B. F. Constraints on asthenospheric flow from the depths of oceanic spreading centers: the East Pacific Rise and the Australian-Antarctic discordance. *Geochem. Geophys. Geosyst.* **10,** Q09007 (2009).
27. Höink, T., Lenardic, A. & Richards, M. Depth-dependent viscosity and mantle stress amplification: implications for the role of the asthenosphere in maintaining plate tectonics. *Geophys. J. Int.* **191,** 30–41 (2012).
28. Parsons, B. & McKenzie, D. Mantle convection and the thermal structure of the plates. *J. Geophys. Res. Solid Earth* **83,** 4485–4496 (1978).
29. Haxby, W. F. & Weissel, J. K. Evidence for small-scale mantle convection from Seasat altimeter data. *J. Geophys. Res. Solid Earth* **91,** 3507–3520 (1986).
30. Richter, F. M. & Parsons, B. On the interaction of two scales of convection in the mantle. *J. Geophys. Res. Solid Earth* **80,** 2529–2541 (1975).
31. French, S., Lekic, V. & Romanowicz, B. Waveform tomography reveals channeled flow at the base of the oceanic asthenosphere. *Science* **342,** 227–230 (2013).
32. Sandwell, D. T., Müller, R. D., Smith, W. H. F., Garcia, E. & Francis, R. New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure. *Science* **346,** 65–67 (2014).
33. Argus, D. F. & Gordon, R. G. No-net-rotation model of current plate velocities incorporating plate motion model NUVEL-1. *Geophys. Res. Lett.* **18,** 2039–2042 (1991).

Acknowledgements We thank the scientific party, captain, crews, and technical teams of RV *Marcus G. Langseth* and RV *Melville* for work that made this study possible. The OBS were provided and supported by Scripps Institution of Oceanography's facility as part of the US Ocean Bottom Seismograph Instrument Pool (http://www.obsip.org). This work was funded by the US National Science Foundation. P.-Y.P.L. thanks the Institute of Earth Science, Academia Sinica, Taipei, Taiwan and Institute of Undersea Technology, National Sun Yat-sen University, Kaohsiung, Taiwan for support during completion of this work.

Author Contributions P.-Y.P.L. and J.B.G. collaborated in developing the concept of this paper and writing the first draft. All authors contributed to the scientific discussion, including presentation of results, interpretation, and implications.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.-Y.P.L. (pylin.patty@gmail.com).

METHODS

Tilt and compliance noise correction for OBS vertical-component data. OBS data often contain substantial noise sources that are not present in land-based seismic records. OBS vertical data are contaminated by compliance noise induced by infragravity waves that are also recorded on bottom-pressure data[34], and tilt noise induced by bottom currents that leaks from horizontal components[35]. We calculate spectral transfer functions between pressure-vertical and horizontal-vertical component data, and apply this transfer function to the vertical components to remove the compliance and tilt noise. At this location and water depth (about 5,200 m), the coherent tilt noise is usually modest, and results in a subtle reduction in noise at periods longer than about 60 s (Extended Data Fig. 1). The compliance signal is quite strong, but is limited to periods longer than about 90 s, and after correction the noise levels of vertical data decrease substantially at long periods. Owing to possible spatial and temporal variation in noise characteristics, the transfer functions of each station are determined for each event individually[36]. To construct the transfer functions, we calculated the spectra and coherence of all pairs of vertical- and horizontal-component seismic data, and differential-pressure data, for two days before each event recording. The two-day-long seismogram was cut into 40 overlapping time windows of length 6,000 s, and spectra, coherence, and transfer functions were calculated and applied as described by refs 1 and 2.

Phase velocity analysis. At long periods (20–150 s), we estimate phase velocities within the array by applying the Automated Surface-Wave Measuring System[37] on Rayleigh waveforms (Extended Data Fig. 2a) of 19 teleseismic earthquakes with moment magnitude over 6.5 (Extended Data Fig. 3). Frequency-dependent phase delays between pairs of nearby stations are measured using cross correlation; the phase delays are then inverted for phase velocities via the Eikonal equation[38]. Lateral variations in phase velocities are small, and we utilize the dispersion in the centre of the array for each event to characterize average structure within the array. The measured phase velocities are strongly dependent on station-to-source azimuth, and at each frequency we solve for the azimuthally averaged phase velocity, plus the magnitude and direction of azimuthal variations of the form $\cos(2\theta) + \cos(4\theta)$, where θ represents azimuth from the centre of the array[39].

We apply the ambient-noise cross-correlation technique to obtain the phase velocities at short periods (10–20 s). The cross-correlation spectra are stacked to retrieve surface-wave Green's functions between station pairs (Extended Data Fig. 2b), and the frequency-dependent phase velocity curve for each station-to-station path is estimated by fitting the cross-spectra[40]. These inter-station dispersion curves are inverted for azimuthally averaged phase velocity plus $\cos(2\theta) + \cos(4\theta)$ variations at each frequency. While we include 4θ terms for the azimuthal variations in both ambient noise and earthquake observations, the amplitude of these terms is very small for all periods (0.09% to 0.4%), and the changes to the 2θ terms is indistinguishable. F-test calculations suggest that these estimates of azimuthal anisotropy do not substantially improve the fit to the data when compared to models including only 2θ variation, so we utilize the 2θ-only estimates to invert for azimuthally anisotropic velocity structure. We use a bootstrap resampling algorithm to test the robustness and stability of the 2θ estimates at each frequency; the error bars on the observations (Fig. 3 and Extended Data Fig. 4) conservatively include the formal errors of the misfit, as well as the potential variance inferred from the bootstrap analysis.

Inversion for azimuthally anisotropic shear velocity. The azimuthally averaged phase velocity dispersion curve is used to estimate a shear-velocity profile as a function of depth via linearized least-squares inversion[41]. We calculate a set of acceptable one-dimensional V_{sv} models using a set of 100 starting models. The 2θ azimuthal dependence can be expressed as combinations of the elastic parameters of B, H and G (ref. 42), with the Rayleigh waves dominated by the parameters for G (ref. 43). In this study, we then modelled the $\cos(2\theta)$ azimuthal variations in phase velocity by using combinations of the depth-dependent parameters G_c and G_s, where the subscripts c and s denote cosine and sine azimuthal dependence[44]. The amplitude and azimuth of G as a function of depth can be computed as $\sqrt{G_c^2 + G_s^2}$ and $0.5\tan^{-1}\frac{G_s}{G_c}$, respectively.

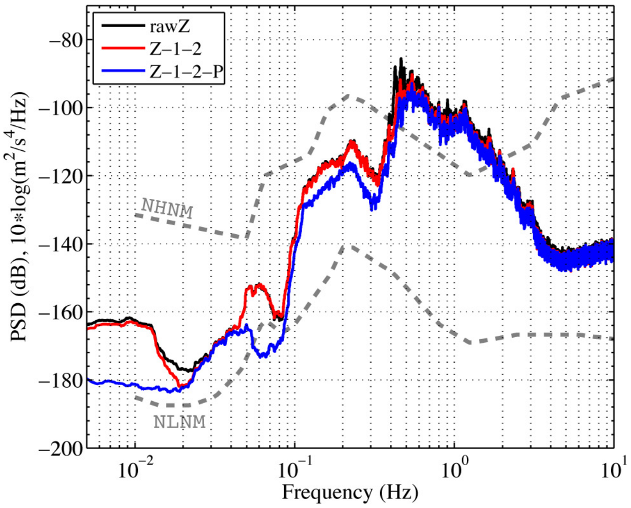
The observed peak-to-peak amplitude A and fast direction ϕ as a function of period can be used to define the data parameters D_c and D_s, where the subscripts c and s denote cosine and sine fast-direction dependence: $D_c = A\cos2\phi$ and $D_s = A\sin2\phi$.

The kernels that relate the data parameters D_c and D_s to model parameters and can be assumed to be the same down to a depth of 500 km as $L = \rho V_{sv}^2$ (ref. 44). We use the average of the best-fitting V_{sv} model determined from the average dispersion curve for the kernel calculation in the azimuthal inversion. We invert for G_c and G_s with a smooth parameterization and a linearized least-squares algorithm that minimizes misfit of D_c and D_s (ref. 42).

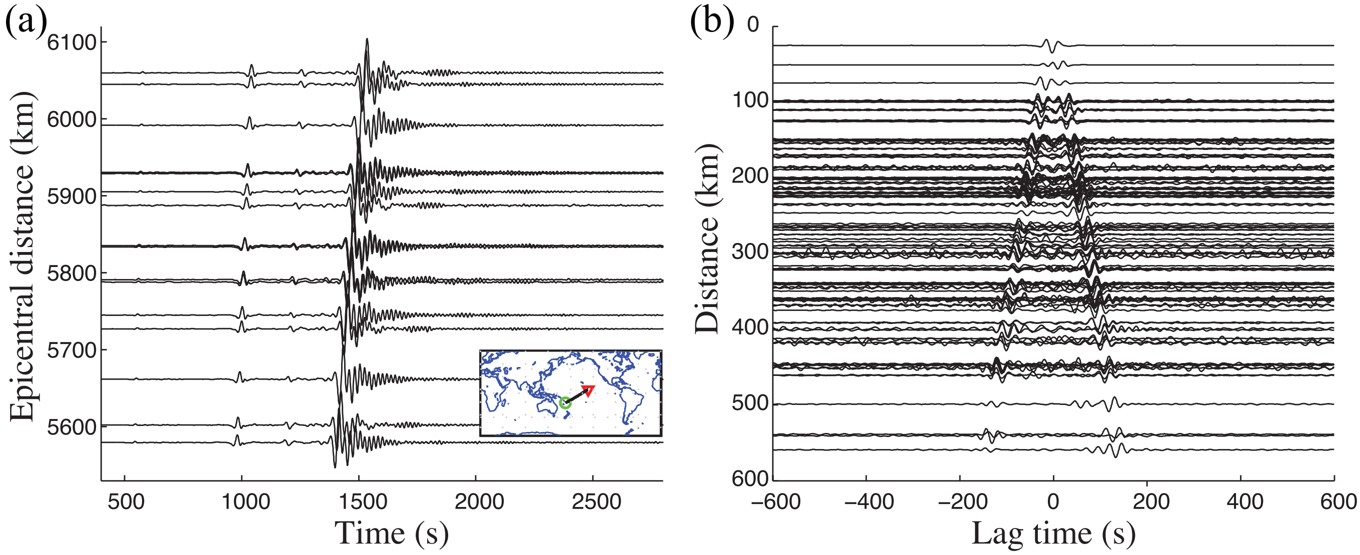
Forward calculations. We utilize forward modelling to explicitly test the robustness of three key characteristics of the inverted anisotropy models (Extended Data Fig. 4). First, we test for the necessity of the secondary peak in anisotropy deep in the asthenosphere. This feature is required to explain the persistent strong azimuthal signal observed in phase velocity at periods of 90 s and longer. Squeezing tests suggest that this structure probably persists to at least 300 km; restricting it to smaller depth maintains a peak in amplitude near 200–250 km, with the amplitude of the peak increasing dramatically as the depth distribution is pushed to depths of <300 km. Second, we test for the possibility of the smoother models for the strength and the direction of the anisotropy (orange solid lines) that specifically do not have a minimum of amplitude in the asthenosphere. Owing to the weakening in the observed strength of the peak-to-peak velocity variations between periods of 30–60 s, the smoother model without a notch in amplitude near a depth of 120 km does not fit the decaying pattern in G amplitude. Third, we evaluate whether a layer of asthenospheric anisotropy parallel to apparent plate motion (APM) might be unresolved by our surface-wave data. Such anisotropy is often inferred at the base of the plate[9,23]. We find that, if present, APM anisotropy must be weak (<1%) and within a layer of thickness <20 km to satisfy the observed fast-direction observations at periods of 40 s and greater within the observational uncertainties.

Code availability. NoMelt data has been archived under network code ZA at the IRIS Data Management Center (http://ds.iris.edu/gmap/ZA?timewindow=2011-2013). Software for array analysis of surface waves is available at https://ds.iris.edu/ds/products/aswms/.

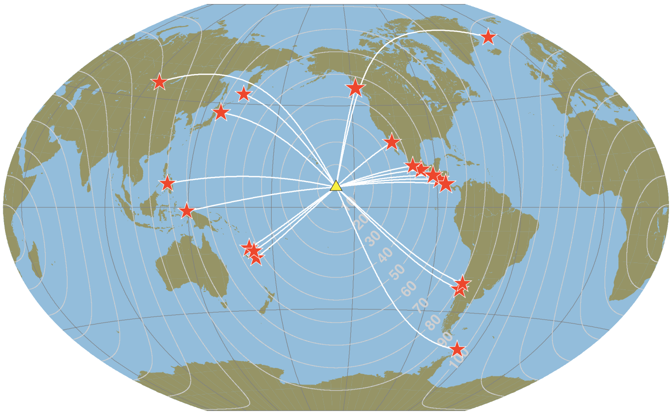
34. Webb, S. C. & Crawford, W. C. Long-period seafloor seismology and deformation under ocean waves. *Bull. Seismol. Soc. Am.* **89,** 1535–1542 (1999).
35. Crawford, W. C. & Webb, S. C. Identifying and removing tilt noise from low-frequency (<0.1 Hz) seafloor vertical seismic data. *Bull. Seismol. Soc. Am.* **90,** 952–963 (2000).
36. Bell, S. W., Forsyth, D. W. & Ruan, Y. Removing noise from the vertical component records of ocean-bottom seismometers: results from year one of the Cascadia Initiative. *Bull. Seismol. Soc. Am.* **105,** 300–313 (2015).
37. Jin, G. & Gaherty, J. B. Surface wave phase-velocity tomography based on multichannel cross-correlation. *Geophys. J. Int.* **201,** 1383–1398 (2015).
38. Lin, F.-C., Ritzwoller, M. H. & Snieder, R. Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophys. J. Int.* **177,** 1091–1110 (2009).
39. Smith, M. L. & Dahlen, F. A. The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium. *J. Geophys. Res. Solid Earth* **78,** 3321–3333 (1973).
40. Menke, W. & Jin, G. Waveform fitting of cross spectra to determine phase velocity using Aki's formula. *Bull. Seismol. Soc. Am.* **105,** 1619–1627 (2015).
41. Herrmann, R. B. Computer programs in seismology: an evolving tool for instruction and research. *Seismol. Res. Lett.* **84,** 1081–1088 (2013).
42. Montagner, J.-P. & Nataf, H.-C. A simple method for inverting the azimuthal anisotropy of surface waves. *J. Geophys. Res. Solid Earth* **91,** 511–520 (1986).
43. Zhu, H. & Tromp, J. Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean. *Science* **341,** 871–875 (2013).
44. Montagner, J. P. & Tanimoto, T. Global upper mantle tomography of seismic velocities and anisotropies. *J. Geophys. Res.* **96,** 20337–20351 (1991).
45. Peterson, J. *Observation and modeling of seismic background noise.* Open-File Report No. 93-322 (US Department of Interior Geological Survey, 1993).



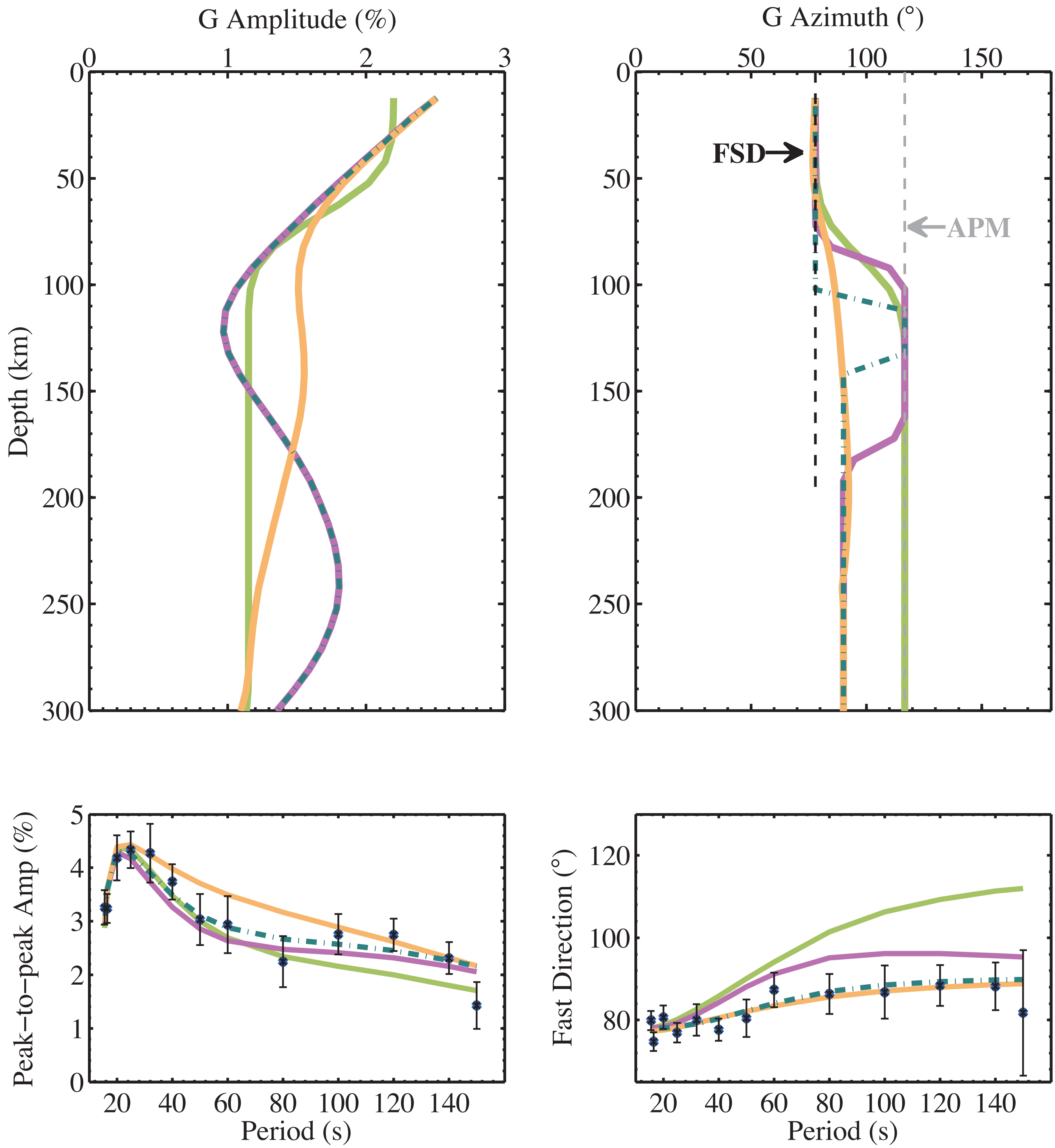
Extended Data Figure 1 | Power spectral densities (PSDs) of raw, tilt and compliance noise-removed seismograms. The PSD values are calculated from vertical data recorded by one NoMelt station, B13, located at 9.25° N, 145.55° W). The black curves show the spectrum of the raw vertical component. The red and blue curves show the spectrum after removing the tilt noise and compliance noise, respectively. The PSD is reported in decibels (dB) referenced to 10log[AAS ($m^{-2}\,s^{-4}\,Hz^{-1}$)], where AAS is the acceleration-amplitude spectrum. The new high- and low-noise models of ref. 45, NHNM and NLNM (thin dashed lines), are shown as references.



Extended Data Figure 2 | Rayleigh waveforms recorded by NoMelt. a, Broadband vertical-component seismograms filtered in the 20–100-s period band from an event (green circle) in the southwestern Pacific (see inset map). The red inverted triangle denotes the NoMelt location. **b**, Ambient noise cross-correlation waveforms, filtered between 10 s and 25 s in the time domain.



Extended Data Figure 3 | Shallow-teleseismic event distribution. 19 events (red stars) with high-quality Rayleigh waves were recorded during NoMelt one-year deployment. White lines represent the corresponding great-circle paths. Grey contours show the epicentral distances at 10° intervals.



Extended Data Figure 4 | Forward calculations. The upper panels show four input models of G parameters (green solid, magenta solid, orange solid and blue dashed-dotted lines). The lower panels show the forward-calculated peak-to-peak amplitude A and fast direction ϕ as functions of period from the four models. Our observations are shown as blue circle symbols with error bars of 1 s.d. as references. The black and grey dashed lines indicate the orientation of the FSD and apparent plate motion (APM) at the NoMelt location.

doi:10.1038/nature18953

A somitic contribution to the apical ectodermal ridge is essential for fin formation

Wouter Masselink[1]†, Nicholas J. Cole[2], Fruzsina Fenyes[1], Silke Berger[1], Carmen Sonntag[1], Alasdair Wood[1], Phong D. Nguyen[1], Naomi Cohen[1], Franziska Knopf[3]†, Gilbert Weidinger[4], Thomas E. Hall[1] & Peter D. Currie[1,5]

The transition from fins to limbs was an important terrestrial adaptation, but how this crucial evolutionary shift arose developmentally is unknown. Current models focus on the distinct roles of the apical ectodermal ridge (AER) and the signalling molecules that it secretes during limb and fin outgrowth. In contrast to the limb AER, the AER of the fin rapidly transitions into the apical fold and in the process shuts off AER-derived signals that stimulate proliferation of the precursors of the appendicular skeleton[1–10]. The differing fates of the AER during fish and tetrapod development have led to the speculation that fin-fold formation was one of the evolutionary hurdles to the AER-dependent expansion of the fin mesenchyme required to generate the increased appendicular structure evident within limbs[11]. Consequently, a heterochronic shift in the AER-to-apical-fold transition has been postulated to be crucial for limb evolution[11]. The ability to test this model has been hampered by a lack of understanding of the mechanisms controlling apical fold induction[11,12]. Here we show that invasion by cells of a newly identified somite-derived lineage into the AER in zebrafish regulates apical fold induction. Ablation of these cells inhibits apical fold formation, prolongs AER activity and increases the amount of fin bud mesenchyme, suggesting that these cells could provide the timing mechanism proposed in Thorogood's clock model of the fin-to-limb transition[11]. We further demonstrate that apical-fold-inducing cells are progressively lost during gnathostome evolution; the absence of such cells within the tetrapod limb suggests that their loss may have been a necessary prelude to the attainment of limb-like structures in Devonian sarcopterygian fish.

To understand the process of fin-fold formation in detail, we systematically dissected how different cell types within the zebrafish fin arise. Muscle and vascular precursors have previously been demonstrated to migrate from the somites to the developing limb bud[13–16]. To define the role that somitic mesoderm contributions have during fin development, we undertook time-lapse analyses using transgenic lines that drive expression of green fluorescent protein (GFP) from *pax3a* regulatory sequences, as *Pax3* expression marks migratory limb myoblasts in amniotes[17]. We determined that somitic mesoderm cells migrate from somites 4–7 to generate myoblasts within the pectoral fin and hypaxial musculature (Fig. 1a–c, Supplementary Video 1). This analysis further identified a subset of the *pax3a*-GFP-positive somite cells within the fin mesenchyme that subsequently migrate and invade the AER (6.4 ± 0.4 (mean ± s.e.m.) cells per fin at 37 h post-fertilization (hpf), $n = 18$). These cells cross the basement membrane that separates the fin mesenchyme from the AER (Fig. 1d–k, Supplementary Video 2), adopting a stereotypical elongated morphology distinct from the underlying cuboidal mesenchymal cells (Fig. 1h–k). Analyses of embryos stained for laminin, which marks the AER basement membrane (Extended Data Fig. 1a–h), as well as histological and transmission electron microscopy studies of the developing fin in the *pax3a*-GFP line (Fig. 1l–o, Extended Data Fig. 1m–p), confirm the position of these cells within the AER. These cells adopt the expression of epithelial

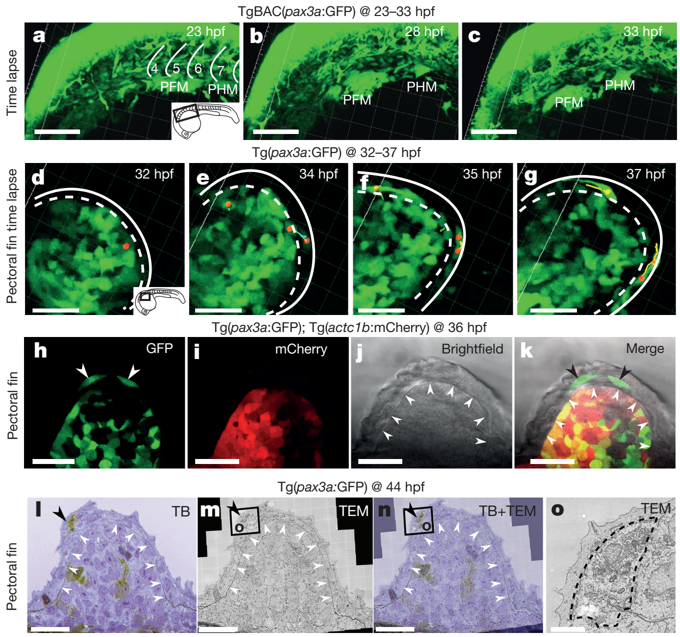


Figure 1 | A dual embryonic origin for zebrafish pectoral fin AER. **a**–**c**, Time-lapse analysis of TgBAC(*pax3a*:GFP) transgenic embryos from 23 to 33 hpf. Somite 4–6 contributes to the zebrafish pectoral fin mesenchyme (PFM), while somite 7 contributes to the posterior hypaxial muscle (PHM) ($n = 6$) (Supplementary Video 1). **d**–**g**, Time-lapse analysis of Tg(*pax3a*:GFP) transgenic embryos from 32–37 hpf. Cells move distally from within the muscle progenitor pool, invade the AER and take on an elongated morphology ($n = 4$). Inserts in **a** and **d** show orientation of images that follow. **h**–**k**, Tg(*pax3a*:GFP); Tg(*actc1b*:mCherry) embryos at 36 hpf identify a subset of GFP-positive mCherry-negative cells with an elongated morphology localized distal to the AER basement membrane (arrowheads, **h** and **k**) ($n = 18$). **j**, **k**, Differential contrast images in which the AER basement membrane can be discerned (white arrowheads). **l**, Toludine blue (TB)-stained section of a 44 hpf Tg(*pax3a*:GFP) embryo stained with an anti-GFP antibody (brown, diaminobenzidine (DAB), black arrowhead) shows the location of this cell distal to the basement membrane (white arrowheads). **m**, Adjacent section scanned with transmission electron microscopy (TEM) reveals that the electron dense DAB-stained cell (black arrowhead) lies distal to the basement membrane (white arrowheads). **n**, Merge of the two adjacent sections. **o**, High-magnification TEM as outlined in **m**. Dashed lines approximate the location of the plasma membrane. Scale bars, 100 μm (**a**–**c**), 20 μm (**d**–**n**) and 3 μm (**o**).

[1]Australian Regenerative Medicine Institute, Level 1, 15 Innovation Walk, Monash University, Wellington Road Clayton, Victoria 3800, Australia. [2]MND and Neurodegenerative Diseases Research Program, Faculty of Medicine and Heath Science, Macquarie University, New South Wales 2109, Australia. [3]The Kennedy Institute of Rheumatology, University of Oxford, Roosevelt Drive, Oxford OX3 7FY, UK. [4]Institute of Biochemistry and Molecular Biology, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm, Germany. [5]EMBL Australia Melbourne Node, Level 1, Building 75, Monash University, Wellington Road, Clayton, Victoria 3800, Australia.
†Present address: DFG- Center for Regenerative Therapies Dresden, Technische Universität Dresden, Fetscherstraße 105, 01307 Dresden, Germany.

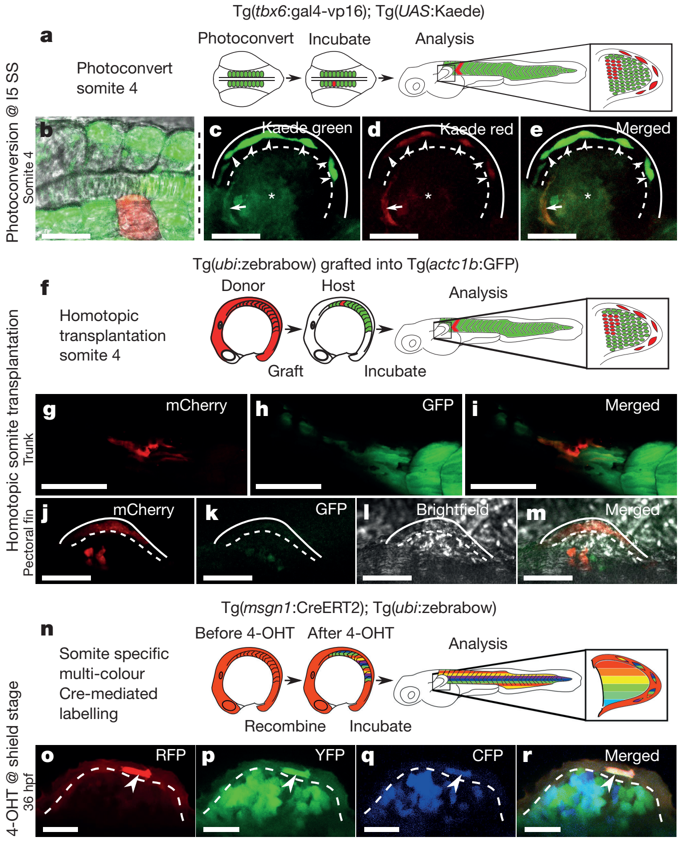


Figure 2 | A specific lineage of somite 4 contributes to the AER in zebrafish. a–e, Photoconversion of somite 4 in embryos transgenic for Tg(*tbx6*:gal4-vp16) and Tg(*UAS*:Kaede) at the 15-somite stage (SS) (**b**), imaged at 2 dpf (**c–e**), reveals photoconverted Kaede (red) within the AER ($n = 11$), (schematised in **a**). **c–e**, A sub-set of the muscle progenitor pool (asterisks), presumptive vasculature (arrows) and cells in the AER are labelled (arrowheads). **f–m**, A Tg(*ubi*:zebrabow) somite (red, **g**, **i**) homotopically transplanted into a host embryo transgenic for Tg(*actc1b*:GFP), which marks muscle in green (**h**, **i**), identifies somite 4 as the source of somitic mesoderm-derived cells in the AER at 48 hpf (**j–m**) (schematised in **f**). **n–r**, Somite-specific 'zebrabow' lineage analysis (schematised in **n**). 4-OHT, 4-hydroxytamoxifen. **o–r**, Tg(*msgn1*:CreERT2); Tg(*ubi*:zebrabow) double-transgenic embryos demonstrate recombined zebrabow expression patterns in the pectoral fin AER at 36 hpf (arrowhead, $n = 4$). Dashed white lines indicate basement membrane. CFP, cyan fluorescent protein; RFP, red fluorescent protein; YFP, yellow fluorescent protein. Scale bars, 50 μm (**b–e**, **j–m**), 100 μm (**g–i**) and 20 μm (**o–r**).

markers within the AER (Extended Data Fig. 1i–l), but do not at any stage express markers of other fin cell types known to be somitically derived such as muscle ($n = 14$, Fig. 1h–k, Extended Data Fig. 2d–g) or vasculature ($n = 12$, Extended Data Fig. 2a–c).

These observations unexpectedly revealed that the AER, which has long been considered a tissue of pure ectodermal origin, has a dual mesodermal and ectodermal composition in zebrafish. To identify the specific somite that contributes cells to the AER, we undertook a series of distinct fate-mapping strategies. First, regional somite-specific photoconversions were performed using embryos transgenic for the somite-specific driver line Tg(*tbx6*:gal4-vp16) and Tg(*UAS*:Kaede), which confirmed somitic contributions to pectoral fin and hypaxial muscle (Extended Data Fig. 3a–h), the vasculature and the AER ($n = 7$, Extended Data Fig. 3i–k). Sequential photoconversions of individual somites determined that somite 4 specifically provides a contribution to the AER (Fig. 2a–e, Extended Data Fig. 3l–q). Homotopic somite transplantations subsequently confirmed that somite 4 has the unique capacity to contribute to the AER (Fig. 2f–m, Extended Data Fig. 4a–m, Supplementary Video 3). However, mechanical removal of somite 4 did not lead to a loss of somite-derived cells in the AER (Extended Data Fig. 4n–r), suggesting flexibility in cell specification mechanisms not solely dependent on the somite of origin. Lastly, we undertook a 'zebrabow' lineage analysis using the somite-specific CreERT2 line, Tg(*msgn1*:creERT2)[18] (Fig. 2n). Using this approach, we determined that a separate lineage of somite cells, distinct from those that generate fin muscle, contributes to the AER ($n = 9$, Fig. 2o–r). Thus, three independent fate-mapping strategies—photoconversion, somite transplantation, and Cre/*loxP* lineage analyses—reveal the existence of a distinct somitic lineage that contributes to the zebrafish AER.

Despite extensive fate-mapping analyses, a somitic contribution to the AER of tetrapods has not been described[14–16,19,20], neither has a somitic contribution to the AER of the caudal fin been observed[7], together suggesting that somite-derived cells probably influence a process specific to the development of paired fins. The formation of the fin fold and the subsequent generation of the larval fin rays, or actinotrichia, within the fold are crucial for paired fin formation[21–23]. Two pieces of evidence suggest that somite cells could be involved in both of these processes. First, somite cells contribute to the fin AER immediately before the AER-to-apical-fold (AF) transition (Extended Data Fig. 1a–h). Second, antibodies to distinct collagen isoforms that visualize actinotrichia development demonstrate that actinotrichia form directly proximal to the somite-derived cells within the AF (Fig. 3a–p, right insets). Furthermore, an antibody raised against collagen type I α1a (Col1a1a), which is known to be involved in actinotrichia formation[22], specifically localizes to the *pax3a*-GFP-positive cells in the AF (Fig. 3q–u, bottom insets), indicating that these cells are probably directly secreting collagen required for actinotrichia formation. Collectively, these analyses suggest that the somite-derived cells present within the AF are actinotrichia-forming cells, a cell type previously proposed to be required for pectoral fin-fold development[21] but which remained unidentified until now.

Next, ablation of somite-derived cells was performed at two different stages of AF formation using two distinct strategies. First, a genetically targeted cell ablation system was used to remove somite-derived cells in the AER immediately before the AF induction. Crossing the Tg(*tbx6*:gal4-vp16) line to the Tg(*UAS-E1b*:Eco.NfsB-mCherry)[c264] strain results in the expression of the nitroreductase enzyme (NfsB) specifically within somite-derived cells, with the highest levels evident in the cells within the AER and exclusively maintained in these cells by 48 hpf (Fig. 4i, Extended Data Figs 3i–k, 5b). Specific ablation of somitic mesoderm-derived cells within the AER occurred after metronidazole (MTZ) addition (92% ablation, $n = 25$, Extended Data Fig. 5), which led to a severe disruption in AF formation ($n = 11$) and a lack of actinotrichia deposition ($n = 5$, Fig. 4c, g), defects not evident in control conditions (Fig. 4a, b, e, f). Fin fold length in treated larvae was reduced to 38% of untreated controls at 2.5 days post-fertilization (dpf) (MTZ alone 66.0 ± 1.5 μm, $n = 5$; NfsB alone 69.8 ± 1.0 μm, $n = 5$; MTZ$^+$/NfsB$^+$ 31.8 ± 0.4 μm, $n = 5$, Fig. 4d), increasing only to 59% of the length of controls at 4 dpf (MTZ alone 225.2 ± 3.7 μm, $n = 5$; NfsB alone 237.0 ± 2.2 μm, $n = 5$; MTZ$^+$/NfsB$^+$ 147.9 ± 0.8 μm, $n = 5$, Fig. 4h). Collectively, these analyses indicate that somite-derived AER cells are required for fin-fold induction and they have therefore been termed apical-fold-inducing cells (AFICs).

The second approach involved the laser ablation of AFICs after migration into the AER. A single cell ablation method was established that used the property of the photoactivatable protein Kaede to absorb laser light efficiently at 405 nm[24] (Extended Data Fig. 6, Supplementary Video 4). Using this method, Kaede-positive cells within the AF of Tg(*tbx6*:gal4-vp16); Tg(*UAS*:Kaede) double-transgenic animals were ablated at 40 hpf, a time point immediately after both fin-fold induction and the initiation of actinotrichia formation (Fig. 3e–h, Extended Data Fig. 1a–h; cells remaining after ablation 0.4 ± 0.24, $n = 5$, Fig. 4j, Extended Data Fig. 7). By contrast, contralateral control fins possessed a normal number of Kaede-positive cells within the distal

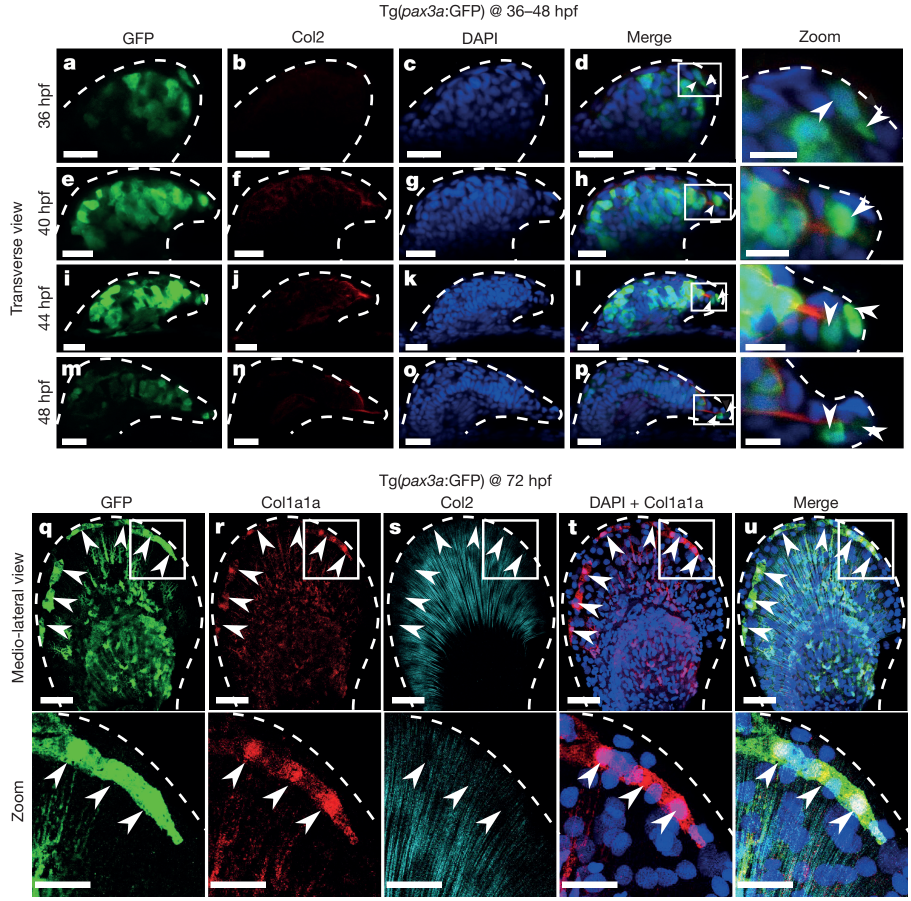


Figure 3 | Somitic mesoderm-derived cells in the AER express actinotrichia-specific collagens. a–p, Transverse views of Tg(*pax3a*:GFP) pectoral fins stained for GFP (green), collagen 2 (Col2; red), and DAPI (blue). Pectoral fins at 36 hpf (**a–d**) ($n=7$), 40 hpf (**e–h**) ($n=8$), 44 hpf (**i–l**) ($n=8$), and 48 hpf (**m–p**) ($n=6$). Actinotrichia growth radiates directly from *pax3a*:GFP-positive cells within the AF at 44 hpf (**j, l**) and 48 hpf (**n, p**). High-magnification views of the boxed region of the extending fin fold are shown in the right panels of **d**, **h**, **l** and **p**. Arrowheads denote *pax3a*:GFP-positive cells within the AF. **q–u**, *pax3a*:GFP-positive cells express Col1a1a. Staining of Tg(*pax3a*:GFP) pectoral fins at 3 dpf for GFP (**q**), Col1a1a (**r**), Col2 (**s**), DAPI and Col1a1a (**t**) and merged images (**u**) ($n=7$). High-magnification views of **q–u** are shown in the bottom panels. Scale bars, 20 μm (**a–p**), 10 μm (**d, h, l, p**, high-magnification, right panels), 50 μm (**q–u**) and 25 μm (**q–u**, high-magnification, bottom panels). Dashed lines indicate the distal fin edge.

fin fold (9.2 ± 0.37, $n=5$, Fig. 4i). Ablation of AFICs from the AF resulted in a notable reduction in the size of actinotrichia, which fail to undergo distal extension after ablation (Fig. 4j), a defect not evident within contralateral, non-ablated fins (Fig. 4i). Furthermore, the fin fold itself was reduced to nearly 30% of its overall length after ablation (contralateral 435.3 ± 5.2 μm, $n=5$; laser-ablated 315.1 ± 8.9 μm, $n=5$, Fig. 4k). By contrast, ablation of cells from the AER ectoderm, before the invasion of AFICs, had no effect on fin development ($n=5$, Extended Data Fig. 7w–zc), indicating that non-specific cell loss from the ectoderm was not responsible for the observed phenotype. Furthermore, in both of our ablation regimes the mesenchyme of the endoskeletal disc underwent a significant expansion (MTZ ablation; MTZ alone 208.2 ± 1.9 μm, $n=5$; NfsB alone 214.1 ± 1.7 μm, $n=5$; MTZ$^+$/NfsB$^+$ 246.0 ± 4.2 μm, $n=5$; laser ablation; non-ablated contralateral fin 389.8 ± 4.5 μm, $n=5$; laser-ablated fin 426.5 ± 3.7 μm, $n=5$, Fig. 4h, k). Collectively, these analyses reveal that ablation of AFICs disrupts both fin-fold induction and actinotrichia formation and leads to an expansion of the underlying fin mesenchyme.

To test whether this increased proliferation is due to continued signalling from the AER to the underlying mesenchyme in ablated larvae, we assessed the gene expression patterns of *fgf24* and *fgf8*, two major proliferative signals in the zebrafish fin[1,25,26]. *fgf24* undergoes a shift in expression from the underlying fin mesenchyme to the AF at the time of fin-fold formation[25], a shift that fails to occur after MTZ treatment, suggesting a delay in AER maturation occurs within ablated larvae (Fig. 4l, m, $n=12$, 75% affected). *fgf8* expression is specifically restricted to the AER until it is reduced during AF expansion (Fig. 4n). In ablated larvae, however, both the levels of expression and the region of the AER that expresses *fgf8* expand markedly when compared to non-ablated controls (Fig. 4n, o, $n=13$, 77% affected). Furthermore, the mitotic index within the mesenchyme of ablated fins was more than double that exhibited by untreated controls (MTZ alone 0.143 ± 0.004, $n=12$; NfsB alone 0.148 ± 0.004, $n=12$; MTZ$^+$/NfsB$^+$ 0.342 ± 0.006, $n=12$; Extended Data Fig. 8a–d), an increase inhibited by the addition of the Fgf signalling inhibitor SU5402 (Extended Data Fig. 8e). These results indicate that the increase in proliferation evident in ablated fins is Fgf dependent, a result concordant with embryological studies that prolonged fin AER signalling and inhibited AF induction through the serial removal and regeneration of the zebrafish AER[10].

To understand better the evolutionary importance of our findings, we examined the existence of AFICs during pectoral fin development within the embryos of species strategically positioned in the gnathostome phylogeny (Fig. 4p–w, Extended Data Fig. 9). Col1a1a immunoreactivity reveals that cells analogous to the AFICs of zebrafish are present within the AER/AF of all species examined, although they vary in number and the timing of residency. Embryos of the basal gnathostome, the epaulette shark (*Hemiscyllium ocellatum*), exhibit a large number of Col1a1a-expressing cells within the AER of their pectoral fins, which persist at all developmental stages examined (Fig. 4p–r, Extended Data Fig. 9a–h). Embryos of the basal

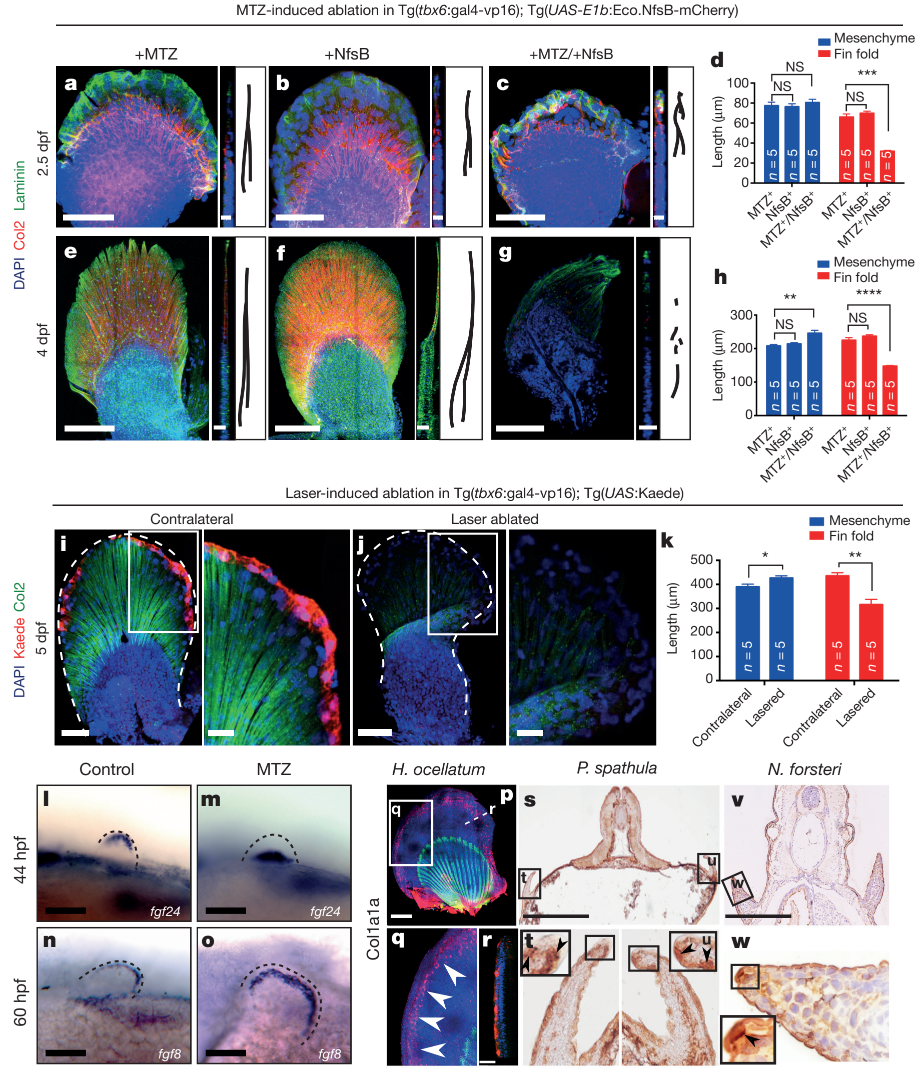


Figure 4 | Somitic mesoderm-derived cells are required for fin-fold induction and actinotrichia formation, and are present within fin AER throughout the gnathostome phylogeny. a–h, Embryos transgenic for Tg(*tbx6*:gal4-vp16) with/without the Tg(*UAS-E1b*:Eco.NfsB-mCherry) transgene were incubated with/without MTZ. Staining for Col2 (red), pan-laminin (green), and DAPI (blue) were performed at 2.5 dpf (**a–c**) ($n = 5$, each group) and 4 dpf (**e–g**) ($n = 5$, each group). Digital sections reveal MTZ-induced AFIC ablation results in a block of fin-fold formation (**c**, **g**, middle panels), schematically represented in an outline of fin-fold shape (**c**, **g**, right panels). NfsB$^-$/MTZ$^+$ (**a**, **e**) and NfsB$^+$/MTZ$^-$ (**b**, **f**) controls do not reveal disruption in fin-fold formation. Fin fold length is reduced and mesenchyme expanded at 2.5 dpf (**d**) and 4 dpf (**h**). **i**, **j**, Laser-mediated ablation of Kaede-positive cells in the AF of Tg(*tbx6*:gal4-vp16); Tg(*UAS*:Kaede) double-transgenic embryos at 40–44 hpf analysed at 5 dpf. Kaede-positive cells are absent in the fin fold of laser-ablated fins (**j**) ($n = 5$), whereas Kaede expression is persistent in contralateral non-ablated control fins (**i**) ($n = 5$) (red). Actinotricha formation is disrupted and morphology is altered (dashed white lines) in laser-ablated fin folds (**j**) compared to the contralateral control (**i**). **k**, Quantification of proximal distal axis measurements in **i** and **j**, revealing a reduction in the size of the ablated fin fold, while the mesenchyme is increased in size. **l–o**, *In-situ* hybridization for *fgf8* and *fgf24* on Tg(*tbx6*:gal4-vp16); Tg(*UAS-E1b*:Eco.NfsB-mCherry) ablated and non-ablated embryos. Dashed lines demarcate the pectoral fin. At 44 hpf, *fgf24* expression localizes to the AER of control animals (**l**) while in MTZ-treated animals expression remains in the underlying mesenchyme (**m**). At 60 hpf *fgf8* expression is maintained throughout the AER of MTZ-treated animals (**o**) compared to control animals (**n**), where it becomes restricted to the distal tip of the fin fold. **p–r**, A fin of an embryo of the epaulette shark *H. ocellatum* at 65-mm stage (approximately *S. canicula* stage 30/31) stained for Col1a1a (red), skeletal muscle (green) and DAPI (nuclear stain; blue). Col1a1a-expressing cells are located throughout the AER at high density. **q**, High-magnification views of boxed region in **p**. **r**, Digital cross section of the dorsal half of the distal most fin edge at the level shown in **p**. **s–u**, A 3 dpf embryo of the basal actinopterygian *P. spathula*, transverse sectioned and stained for Col1a1a expression (brown, DAB). **t**, **u**, High-magnification views of the boxed regions in **s**. Arrows in inset indicate Col1a1a-expressing cells. **v**, **w**, *N. forsteri* (Australian lung fish) stage 47 embryo in transverse section and stained for Col1a1a and counterstained with haematoxylin and eosin. A single Col1a1a-expressing cell is evident in the most distal tip of the AER. Scale bars, 50 µm (**a–c**, **l–o**), 10 µm (**a–c**, high-magnification, middle panels), 100 µm (**e–g**, **i**, **j**), 25 µm (**e–g**, high-magnification, middle panels; **i**, **j**, high-magnification, right panels), 400 µm (**p**) and 500 µm (**s**, **v**). $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$ (analysis of variance (ANOVA) with Tukey's post-hoc analysis (**d**, **h**), and unpaired Student's two-tailed *t*-test (**k**)). NS, not significant; error bars show s.e.m.

actinopterygian, the North American paddlefish (*Polyodon spathula*), possess Col1a1a-expressing cells within the AER/AF at similar numbers and timing to that evident in zebrafish, but reduced in number in comparison to shark (Fig. 4s–u, Extended Data Fig. 9i–p). Embryos of the Australian lungfish (*Neoceratodus forsteri*), a sarcopterygian fish species that phylogenetically abuts the tetrapod transition, possess the least number of Col1a1a-expressing cells within the AER of their pectoral fins, which are lost as the fin-fold develops (Fig. 4v–w, Extended Data Fig. 9q–x). The results of sampling these key phylogenetic species, combined with known data on limb development

in tetrapods[14–16,19,20], are broadly consistent with a model that suggests a reduction in AFICs over evolutionary time would allow for a greater expansion of the mesenchyme evident in tetrapod species. Our results, therefore, provide a mechanistic basis to explain how a heterochronic shift in AER activity could have been achieved developmentally (Extended Data Fig. 10), and in doing so provides experimental support for Thorogood's theoretical clock model of fin to limb transition[11].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 July 2015; accepted 20 June 2016.
Published online 20 July 2016.

1. Grandel, H., Draper, B. W. & Schulte-Merker, S. dackel acts in the ectoderm of the zebrafish pectoral fin bud to maintain AER signaling. *Development* **127,** 4169–4178 (2000).
2. Géraudie, J. Initiation of the actinotrichial development in the early fin bud of the fish, Salmo. *J. Morphol.* **151,** 353–361 (1977).
3. Saunders, J. W. Jr The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *J. Exp. Zool* **108,** 363–403 (1948).
4. Todt, W. L. & Fallon, J. F. Development of the apical ectodermal ridge in the chick wing bud. *J. Embryol. Exp. Morphol.* **80,** 21–41 (1984).
5. Fallon, J. F. & Kelley, R. O. Ultrastructural analysis of the apical ectodermal ridge during vertebrate limb morphogenesis. *Development* **51,** 241–256 (1977).
6. Wood, A. Early pectoral fin development and morphogenesis of the apical ectodermal ridge in the killifish, *Aphyosemion scheeli. Anat. Rec.* **204,** 349–356 (1982).
7. Dane, P. J. & Tucker, J. B. Modulation of epidermal cell shaping and extracellular matrix during caudal fin morphogenesis in the zebra fish *Brachydanio rerio*. *J. Embryol. Exp. Morphol.* **87,** 145–161 (1985).
8. Jurand, A. Ultrastructural aspects of early development of the fore-limb buds in the chick and the mouse. *Proc. R. Soc. Lond. B Biol. Sci.* **162,** 387–405 (1965).
9. Grandel, H. & Schulte-Merker, S. The development of the paired fins in the zebrafish (*Danio rerio*). *Mech. Dev.* **79,** 99–120 (1998).
10. Yano, T., Abe, G., Yokoyama, H., Kawakami, K. & Tamura, K. Mechanism of pectoral fin outgrowth in zebrafish development. *Development* **139,** 2916–2925 (2012).
11. Thorogood, P. in *Developmental Patterning of the Vertebrate Limb* (eds Hinchliffe, J. R. *et al*.) 347–354 (Springer US, 1991).
12. Yano, T. & Tamura, K. The making of differences between fins and limbs. *J. Anat.* **222,** 100–113 (2013).
13. Neyt, C. *et al.* Evolutionary origins of vertebrate appendicular muscle. *Nature* **408,** 82–86 (2000).
14. Christ, B., Jacob, H. J. & Jacob, M. Experimental analysis of the origin of the wing musculature in avian embryos. *Anat. Embryol. (Berl.)* **150,** 171–186 (1977).
15. Chevallier, A., Kieny, M. & Mauger, A. Limb-somite relationship: origin of the limb musculature. *J. Embryol. Exp. Morphol.* **41,** 245–258 (1977).
16. Mayeuf-Louchart, A. *et al.* Notch regulation of myogenic versus endothelial fates of cells that migrate from the somite to the limb. *Proc. Natl Acad. Sci. USA* **111,** 8844–8849 (2014).
17. Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R. & Gruss, P. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J.* **10,** 1135–1147 (1991).
18. Gurevich, D. B. *et al.* Asymmetric division of clonal muscle stem cells coordinates muscle regeneration *in vivo*. *Science* http://dx.doi.org/10.1126/science.aad9969 (2016).
19. Kimmel, R. A. *et al.* Two lineage boundaries coordinate vertebrate apical ectodermal ridge formation *Genes Dev.* **14,** 1377–1389 (2000).
20. Altabef, M., Clarke, J. D. & Tickle, C. Dorso-ventral ectodermal compartments and origin of apical ectodermal ridge in developing chick limb. *Development* **124,** 4547–4556 (1997).
21. Zhang, J. *et al.* Loss of fish actinotrichia proteins and the fin-to-limb transition. *Nature* **466,** 234–237 (2010).
22. Durán, I., Marí-Beffa, M., Santamaría, J. A., Becerra, J. & Santos-Ruiz, L. Actinotrichia collagens and their role in fin formation. *Dev. Biol.* **354,** 160–172 (2011).
23. Webb, A. E. *et al.* Laminin α5 is essential for the formation of the zebrafish fins. *Dev. Biol.* **311,** 369–382 (2007).
24. Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H. & Miyawaki, A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl Acad. Sci. USA* **99,** 12651–12656 (2002).
25. Fischer, S., Draper, B. W. & Neumann, C. J. The zebrafish *fgf24* mutant identifies an additional level of Fgf signaling involved in vertebrate forelimb initiation. *Development* **130,** 3515–3524 (2003).
26. Cohn, M. J., Izpisúa-Belmonte, J. C., Abud, H., Heath, J. K. & Tickle, C. Fibroblast growth factors induce additional limb development from the flank of chick embryos. *Cell* **80,** 739–746 (1995).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors acknowledge E. McGlinn for expert advice, C. Wicking for comments on the manuscript, E. Tanaka for advice and technical support, Fishcore for expert zebrafish care, M. Morsch, V. Oorschot and the Monash electron microscopy facility for expert technical advice, Monash Micro Imaging for microscope availability and maintenance, S. Chow for technical assistance, M. Voz for reagents, T. Carney for fish lines and J. Joss for lungfish rearing and expertise. P.D.C. is supported by an NHMRC Principal Research Fellowship. N.J.C. is supported by The Snow Foundation and BitFury.org. This work was supported by an ARC Discovery Grant DP110101482. The Australian Regenerative Medicine Institute is supported by funds from the State Government of Victoria and the Australian Federal Government.

Author Contributions W.M. designed and performed experiments, analysed data and co-wrote the manuscript, N.J.C. designed and performed experiments, P.D.N., T.E.H., F.K. and G.W. generated reagents and fish strains, S.B., F.F., C.S., A.W. and N.C. performed experiments, P.D.C. designed and performed experiments, analysed data and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.D.C. (peter.currie@monash.edu).

Reviewer Information *Nature* thanks C. Tabin and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Fish husbandry and handling. Adult zebrafish were maintained on a 14-h light/10-h dark cycle using standard procedures[27]. Embryos were maintained at 28 °C in E3 medium. A breeding colony of epaulette sharks (*H. ocellatum*) was established from wild caught specimens in a purpose circulating marine system. Females lay fertilised eggs at a rate of approximately two paired eggs every 2 weeks, and these were removed from breeding tanks and maintained at 26 °C in separate tanks until the desired developmental stage was determined by candling. Formal staging was in length (mm) and with reference to previously published shark developmental staging series[28]. Lungfish (*N. forsteri*) and paddlefish (*P. spathula*) embryos were obtained and staged as previously described[29]. Experimental procedures were approved by the Monash Animal Services Animal Ethics committee, Monash University.

Statistical analysis. No statistical methods were used to predetermine sample size. In each experiment, embryos were randomly selected, and experimenters were blinded to the treatment group prior to analysis. Statistical analysis was performed with a minimum of 5 independent biological replicates; exact numbers have been indicated at relevant locations in the manuscript. Statistical analysis was performed with Prism (GraphPad) using Student's unpaired two-tailed *t*-test when comparing two conditions. Error bars represent s.e.m. Analysis of variance (ANOVA) with Tukey's post-hoc analysis was used when comparing multiple samples.

Zebrafish transgenic lines. Transgenic lines used in this work were: Tg(*actc1b*:GFP)[zf13] (ref. 30), Tg(*kdr1*:mCherry)[ci5] (ref. 31), Tg(*UAS*:Eco.NfsB-mCherry)[c264] (ref. 32), TgBAC(*pax3a*:GFP)[i150] (ref. 33), Tg(*pax3a*:GFP)[pc7], Tg(*tbx6*:gal4-vp16)[sq13] (ref. 34), Tg(*UAS*:Kaede)[rk8] (ref. 35), Tg(*msgn1*:Cre-ERT2)[pc9] (ref. 36), Tg(*ubi*:zebrabow)[a131] (ref. 37), Tg(*act1b*:mCherry)[pc4] (ref. 29), and Tg(*h2a f/z*:nlsKikGR1). Tg(*h2a f/z*:nlsKikGR1) was generated by cloning a 2-kb region of the zebrafish histone 2a promoter[38] next to the nuclear humanized version of the photoconvertible protein KikGR1 (ref. 39) (MBL Corporation). Transgenesis was performed using the Tol2 system[40].

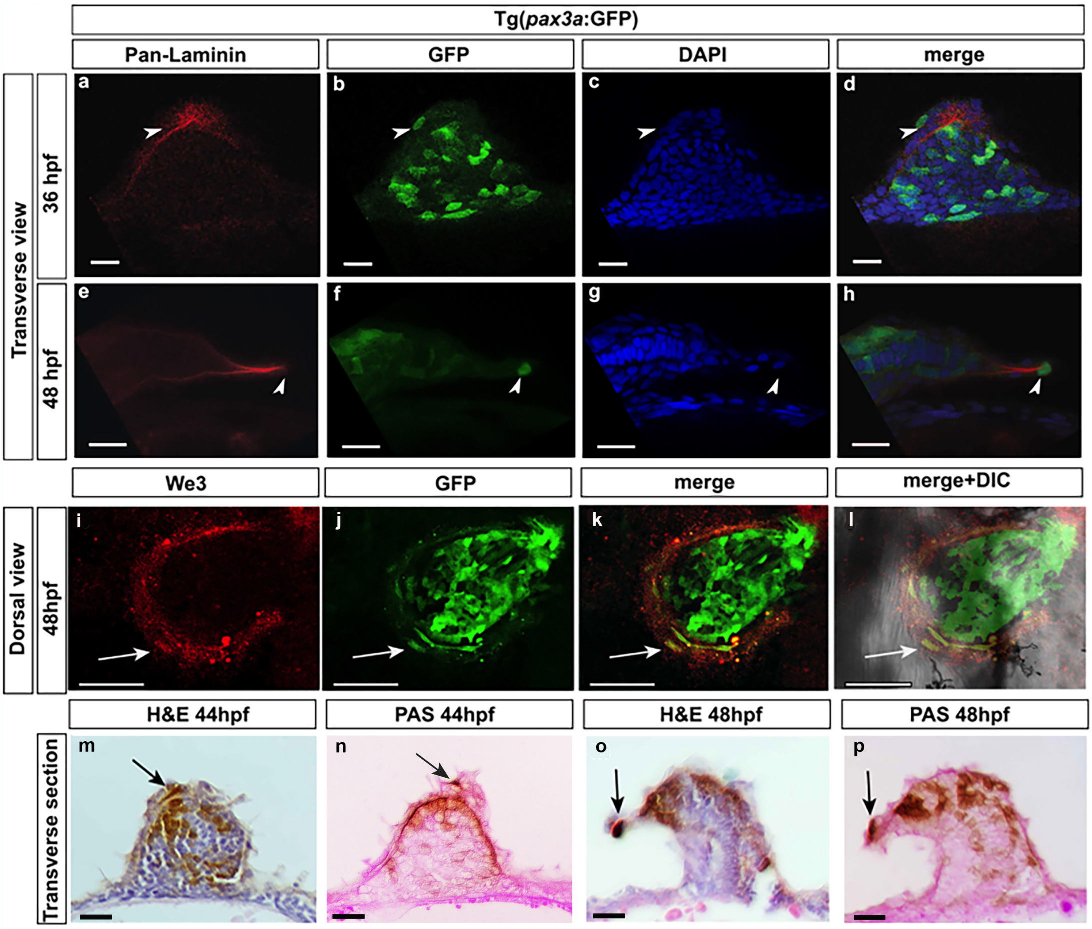
Histology and microscopy. Histological and sectioning methods for haematoxylin and eosin staining, Toluidine blue staining, TEM, and periodic acid-Schiff staining were based on previously published methods[41]. Confocal live imaging was performed on a Zeiss LSM5/710 meta or on a Leica SP5 using a ×20 1.0 N/A water immersion objective. Non-confocal live images were captured on a Zeiss LumarV.12. Antibody-stained samples were imaged on a Leica SP8 using a ×63 1.4 N/A oil objective. Electron micrographs of ultrathin sections were taken on a Nova Nano SEM electron microscope (FEI).

Ablation, photoconversion, transplantation and lineage analysis strategies. Single-cell ablations were performed on a Zeiss LSM5/710 meta using a standard 405-nm diode laser. Using the ROI tool in Zen black 2012 (Zeiss) a region of interest comprising a single Kaede-positive cell was ablated[24]. Single-cell ablations in control fins were performed in a similar fashion, by irradiation of individual Kikume-positive nuclei[39]. The genetically targeted cell-ablation system used the property of the nitroreductase enzyme (Eco.NfsB) to convert the innocuous prodrug MTZ into a cytotoxic agent. To undertake ablation we crossed the Tg(*tbx6*:gal4-vp16) transgenic line to the Tg(*UAS-E1b*:Eco.NfsB-mCherry)[c264] strain, which results in the expression of the nitroreductase enzyme specifically within somite-derived cells. Embryos were incubated with a range of MTZ concentrations at 30 hpf, before fin fold induction. This analysis determined that 20 mM MTZ was the optimal concentration in terms of larval survival (92%, $n = 25$). Nfsb/MTZ ablations were consequently performed using 20 mM MTZ by bathing the embryos from 30–48 hpf. Photoconversion was performed by illuminating selected somites of a 15-somite-stage embryo using the 405-nm diode laser of the Zeiss LSM5/710 meta. Somite transplantations were performed as described previously[29]. Pectoral fin measurements were performed by flat mounting the mechanically dissected pectoral fin. The fin bud mesenchyme was defined using the DAPI signal, which marks the nuclei of these cells. Fin fold length was determined by measuring the distance from the proximal most location of the actinotrichia to the distal most tip of the pectoral fin. Somite extirpation was performed in a similar manner to somite transplantations, without grafting a somite. Cre/*loxP* lineage analysis was performed as described previously[36,37]. Fgf inhibition used SU5402 (Sigma) treatment as described previously[42].

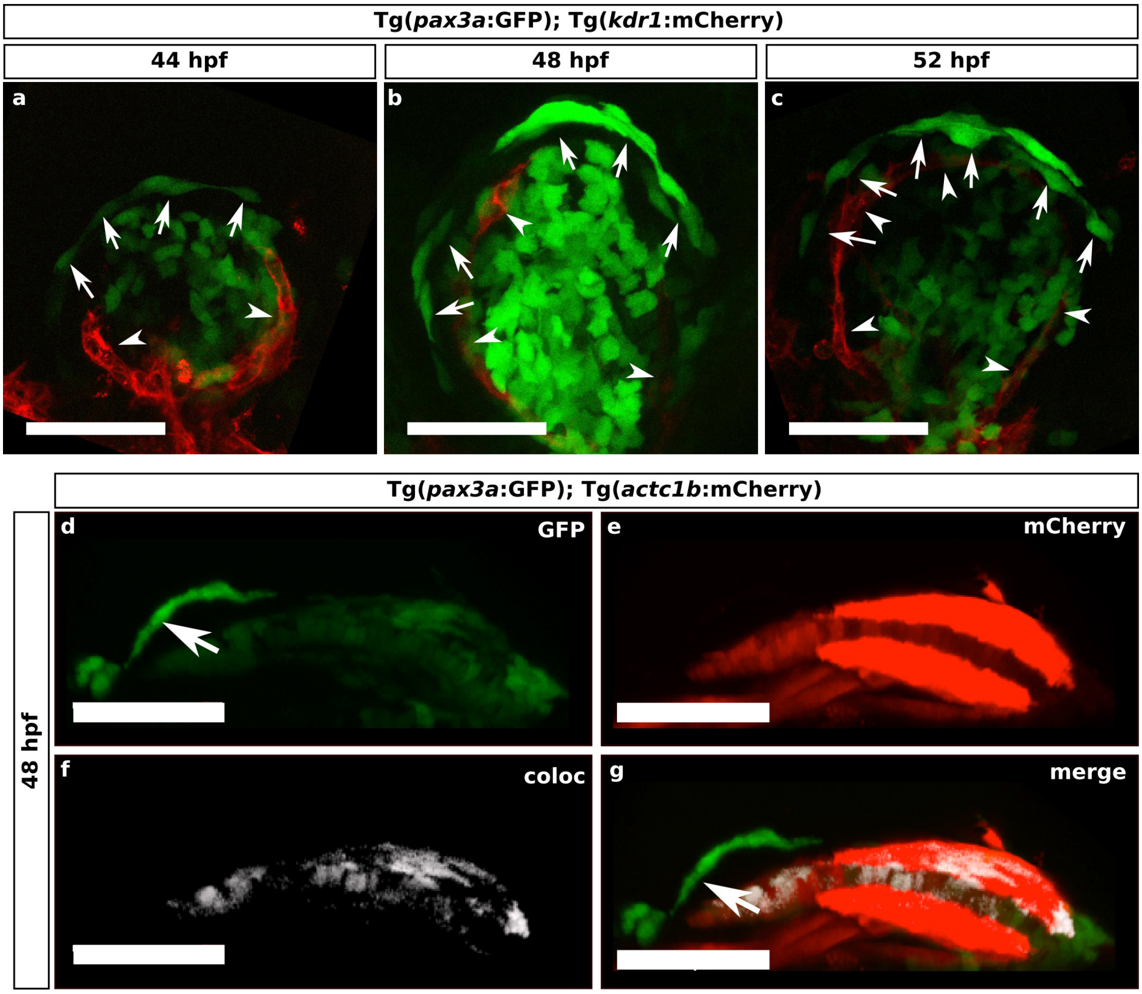
Antibody staining and *in-situ* hybridization. Antibody staining was performed as previously described[27] using antibodies anti-Col2 (DSHB, II-II6B3), anti-Col1a1a (Sigma, SAB2701691), anti-Col1 (Abcam, AB23730), anti-pan-laminin (Sigma, L9393), anti-Kaede (MBL International, PM012), anti-We3 (DSHB, deposited by R. A. Tassava), anti-RFP (Abcam, ab65856), anti-GFP (Life Technologies, A10262), and secondary Alexa Fluor-coupled antibodies (Invitrogen). *In situ* hybridization was conducted as described[43]. Probes used were *fgf8* (Addgene, fb73a06) and *fgf24* (ref. 25).

EdU labelling. Embryos were pulsed with 2.5 mM EdU (Sigma), 10% DMSO for 30 min on ice, followed by 1 h at 28 °C. EdU was visualized according to manufacturer's instructions using Click-iT Chemistry. Mitotic index was determined by dividing the number of EdU-positive cells by the total number of cells.

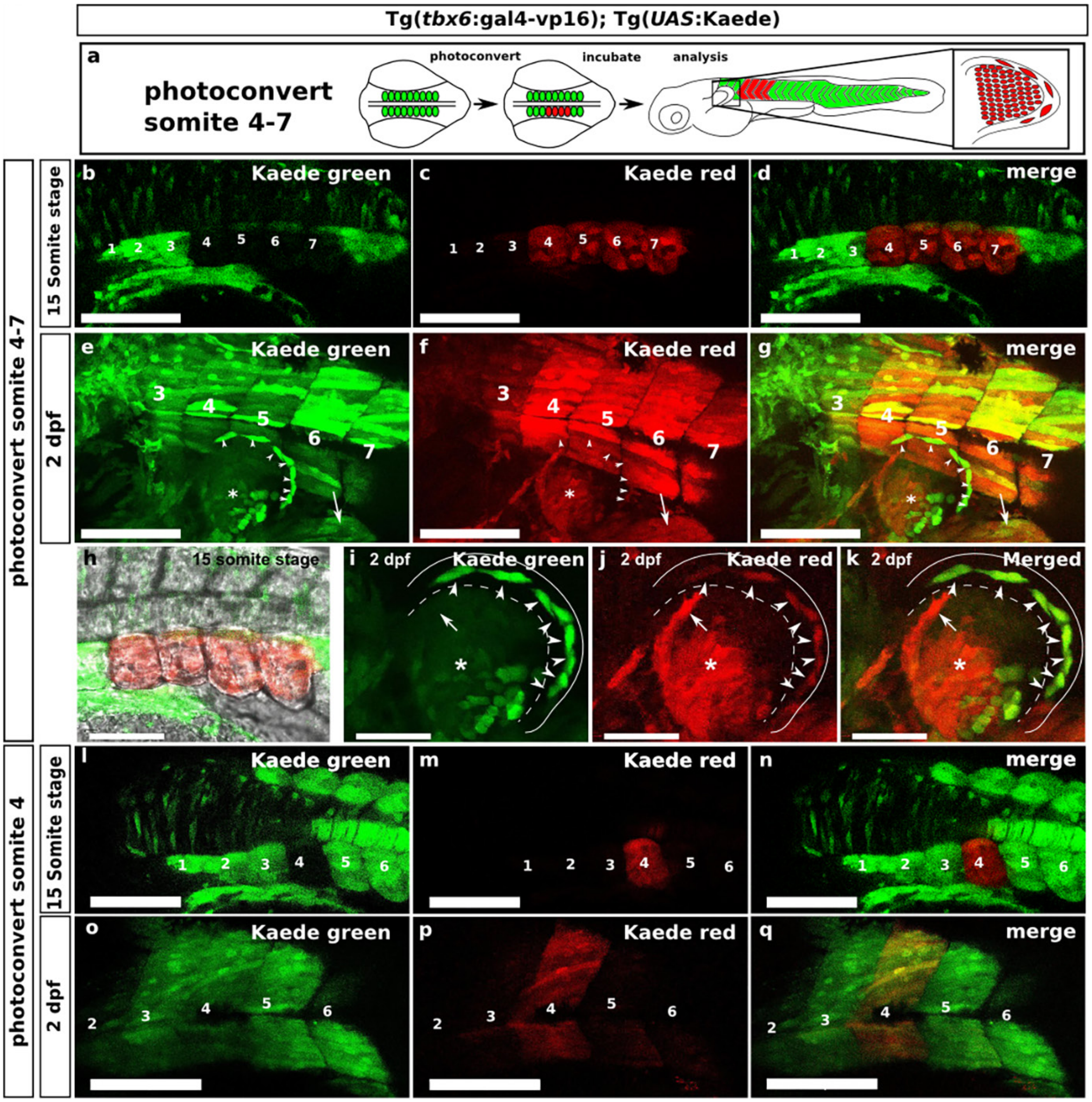
27. Westerfield, M. *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish* (Danio rerio) 4th edn (Univ. Oregon Press, 2000).
28. Ballard, W. W., Mellinger, J. & Lechenault, H. A series of normal stages for development of scyliorhinus-canicula, the lesser spotted dogfish (*Chondrichthyes, scyliorhinidae*). *J. Exp. Zool.* **267,** 318–336 (1993).
29. Cole, N. J. *et al.* Development and evolution of the muscles of the pelvic fin. *PLoS Biol.* **9,** e1001168 (2011).
30. Higashijima, S., Okamoto, H., Ueno, N., Hotta, Y. & Eguchi, G. High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin. *Dev. Biol.* **192,** 289–299 (1997).
31. Proulx, K., Lu, A. & Sumanas, S. Cranial vasculature in zebrafish forms by angioblast cluster-derived angiogenesis. *Dev. Biol.* **348,** 34–46 (2010).
32. Pisharath, H., Rhee, J. M., Swanson, M. A., Leach, S. D. & Parsons, M. J. Targeted ablation of beta cells in the embryonic zebrafish pancreas using *E. coli* nitroreductase. *Mech. Dev.* **124,** 218–229 (2007).
33. Seger, C. *et al.* Analysis of Pax7 expressing myogenic cells in zebrafish muscle development, injury, and models of disease. *Dev. Dyn.* **240,** 2440–2451 (2011).
34. Lee, R. T. H., Knapik, E. W., Thiery, J. P. & Carney, T. J. An exclusively mesodermal origin of fin mesenchyme demonstrates that zebrafish trunk neural crest does not generate ectomesenchyme. *Development* **140,** 2923–2932 (2013).
35. Hatta, K., Tsujii, H. & Omura, T. Cell tracking using a photoconvertible fluorescent protein. *Nat. Protocols* **1,** 960–967 (2006).
36. Nguyen, P. D. *et al.* Haematopoietic stem cell induction by somite-derived endothelial cells controlled by *meox1*. *Nature* **512,** 314–318 (2014).
37. Pan, Y. A. *et al.* Zebrabow: multispectral cell labeling for cell tracing and lineage analysis in zebrafish. *Development* **140,** 2835–2846 (2013).
38. Pauls, S., Geldmacher-Voss, B. & Campos-Ortega, J. A. A zebrafish histone variant H2A.F/Z and a transgenic H2A.F/Z:GFP fusion protein for in vivo studies of embryonic development. *Dev. Genes Evol.* **211,** 603–610 (2001).
39. Habuchi, S., Tsutsui, H., Kochaniak, A. B., Miyawaki, A. & van Oijen, A. M. mKikGR, a monomeric photoswitchable fluorescent protein. *PLoS One* **3,** e3944–e3949 (2008).
40. Fisher, S. *et al.* Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat. Protocols* **1,** 1297–1305 (2006).
41. Berger, J. *et al.* Loss of Tropomodulin4 in the zebrafish mutant trage causes cytoplasmic rod formation and muscle weakness reminiscent of nemaline myopathy. *Dis. Model Mech.* **7,** 1407–1415 (2014).
42. Nguyen-Chi, M. E. *et al.* Morphogenesis and cell fate determination within the adaxial cell equivalence group of the zebrafish myotome. *PLoS Genet.* **8,** e1003014–e1003016 (2012).
43. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protocols* **3,** 59–69 (2008).



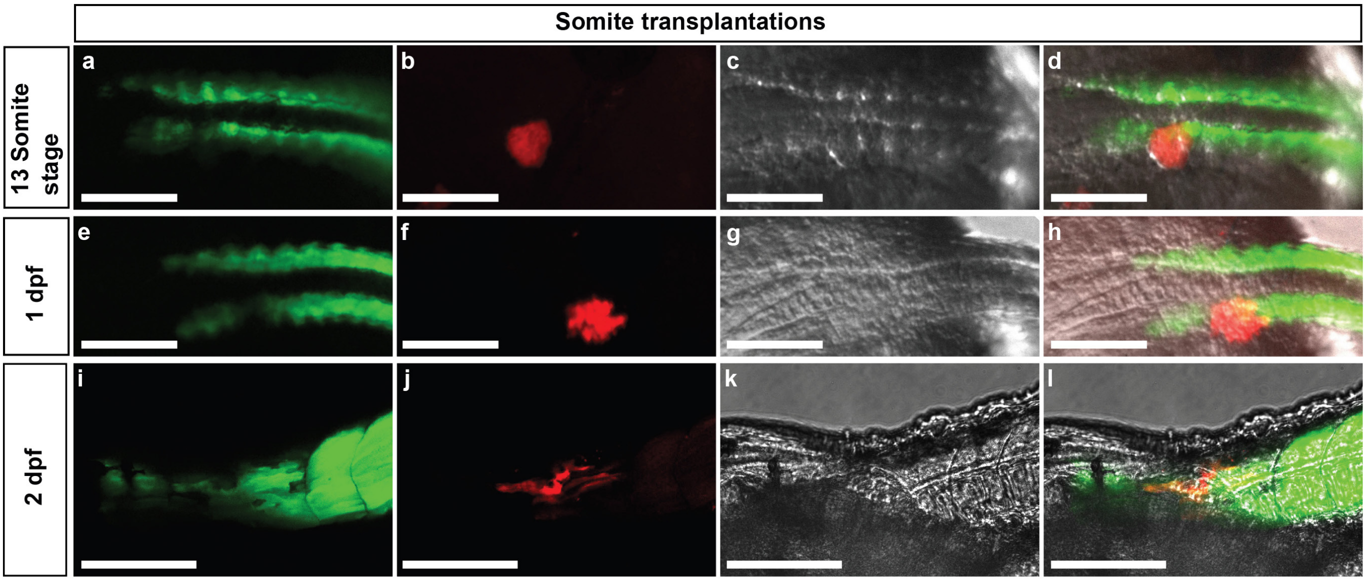
Extended Data Figure 1 | Somitic mesoderm-derived cells migrate into the AER before AF formation and occupy a position distal to the AER basement membrane. a–h, Somite-derived cells are positioned distally to the basement membrane of the AER. Pectoral fin of Tg(*pax3a*:GFP) transgenic embryos at 36 hpf (**a–d**, lateral view) and 48 hpf (**e–h**, transverse view), stained for pan-laminin (red), which marks the basement membrane of the AER (**a**, **e**), GFP (green) (**b**, **f**), DAPI (blue) (**c**, **g**), and a merged view (**d**, **h**). Arrowheads mark the somite-derived cells that invade the AER, and occupy a position distal to the basement membrane. **i–l**, Somite-derived cells are positioned within the AER and adopt expression of an epithelial marker. Dorsal view of a pectoral fin of a Tg(*pax3a*:GFP) embryo at 48 hpf, stained with an antibody against We3 (red), which is an epithelial marker for the AER and wound epithelium in salamander limbs (**i**), GFP (**j**), a merged view (**k**), and a merged differential contrast (DIC) image (**l**). Somite-derived cells (arrows) not only reside in the epithelial AER, but also co-express the epithelial marker. **m–p**, Histological analysis of the positioning of somite-derived cells within the AER and AF. Haematoxylin and eosin (H&E) staining (**m**, **o**) and periodic acid-Schiff (PAS) staining (**n**, **p**) on sagittal sections of Tg(*pax3a*:GFP) pectoral fins at 44 hpf (**m**, **n**) and 48 hpf (**o**, **p**). PAS specifically stains the basement membranes of tissues owing to the presence of Schiff-reactive glycoproteins within these structures. Arrows mark somitic mesoderm-derived cells in the AER (**m**, **n**) and AF (**o**, **p**), respectively. Both these histological analyses reveal that somite-derived cells are positioned distally to a continuous basement membrane. Scale bars, 20 μm.



Extended Data Figure 2 | Distally located GFP-positive cells in the AER of Tg(*pax3a*:GFP) transgenic larvae do not contribute to either vasculature or musculature. a–c, Tg(*pax3a*:GFP) does not label vasculature in the zebrafish pectoral fin. Medio-lateral view of Tg(*pax3a*:GFP); Tg(*kdr1*:mCherry) pectoral fins at 44 hpf (**a**), 48 hpf (**b**), and 52 hpf (**c**). Elongated GFP-positive cells localize distally to the mCherry-positive cells of the developing vasculature (**d**–**g**). Tg(*pax3a*:GFP)-expressing cells in the AER do not colocalize with Tg(*actc1b*:mCherry) expression (**d**). Dorsal and ventral muscle masses are defined by mCherry expression (**e**). GFP and mCherry colocalization shown in white (**f**). Merged image (**g**). Arrows mark GFP-positive cells in the AER, while arrowheads mark the vasculature. Scale bars, 50 μm.

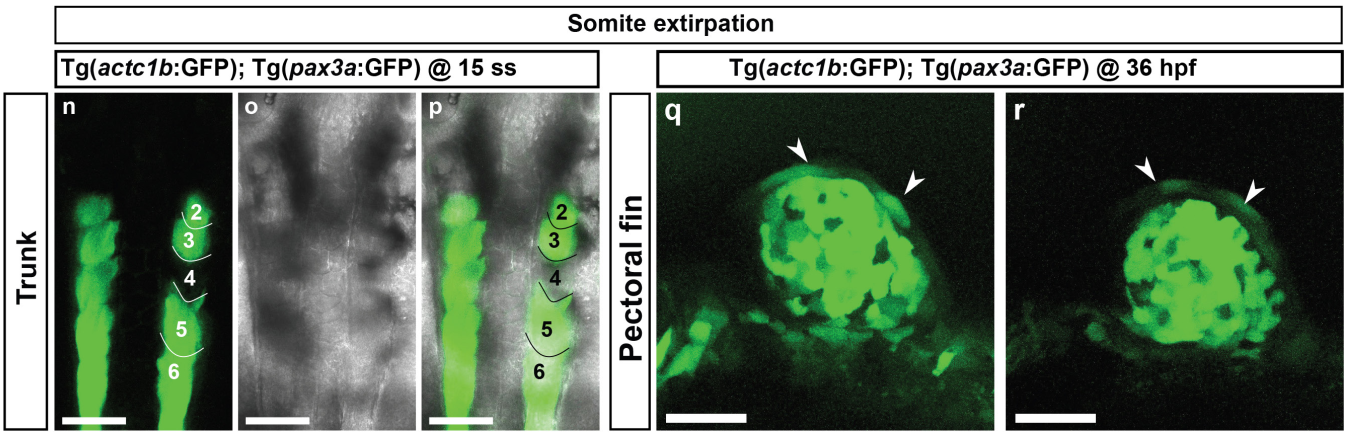


Extended Data Figure 3 | Validation of somite specific photoconversion of Tg(*tbx6*:gal4-vp16); Tg(*UAS*:Kaede) double-transgenic embryos. **a**–**q**, Photoconversion of somite 4–7 (**a**–**k**) and somite 4 (**l**–**q**). **a**, Schematic representation of photoconversion. **b**–**d**, Dorsal views after photoconversion of somite 4–7 at the 15 somite stage reveals non-photoconverted Kaede (green) (**b**) and photoconverted Kaede (red) (**c**). Merged image shown in **d**. Somites 1–7 are numbered. **e**–**g**, Lateral views of the same photoconverted embryo at 2 dpf, revealing non-photoconverted Kaede (**e**) and a photoconverted Kaede (**f**). Merged image shown in **g**. **h**, Close-up view of photoconverted somites from **e**–**g** including brightfield. **i**–**k**, Close-up view of pectoral fin shown in **e**–**g** showing non-photoconverted Kaede (**i**) and photoconverted Kaede (**j**). Merged image shown in **k**. Asterisks indicate presumptive fin muscle progenitors, arrows mark photoconverted Kaede-positive cells of the PHM (**e**–**g**) and presumptive vasculature (**i**–**k**), arrowheads indicate photoconverted Kaede-positive cells in the AER. **l**–**n**, Dorsal view after photoconversion of somite 4 at 15 somites, showing non-photoconverted Kaede (**l**), photoconverted Kaede (**m**), and merged image (**n**). Somites 1–6 are numbered. **o**–**q**, Lateral view of the same photoconverted embryo at 2 dpf, showing non-photoconverted Kaede (**o**), photoconverted Kaede specifically in somite 4 (**p**), and a merged image (**q**). Somites 2–6 are numbered. Scale bars, 100 μm (**b**–**g**, **i**–**q**), 50 μm (**h**–**k**).



m

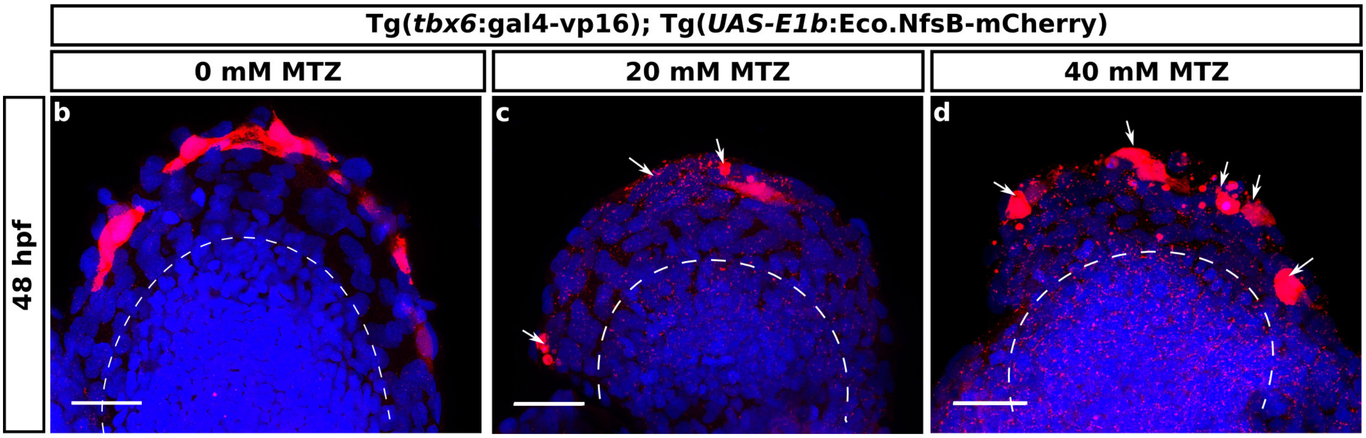
Somite #	n	Pectoral fin contribution muscle	AER
3	4	0	0
4	2	2	2
5	3	3	0
6	2	2	0



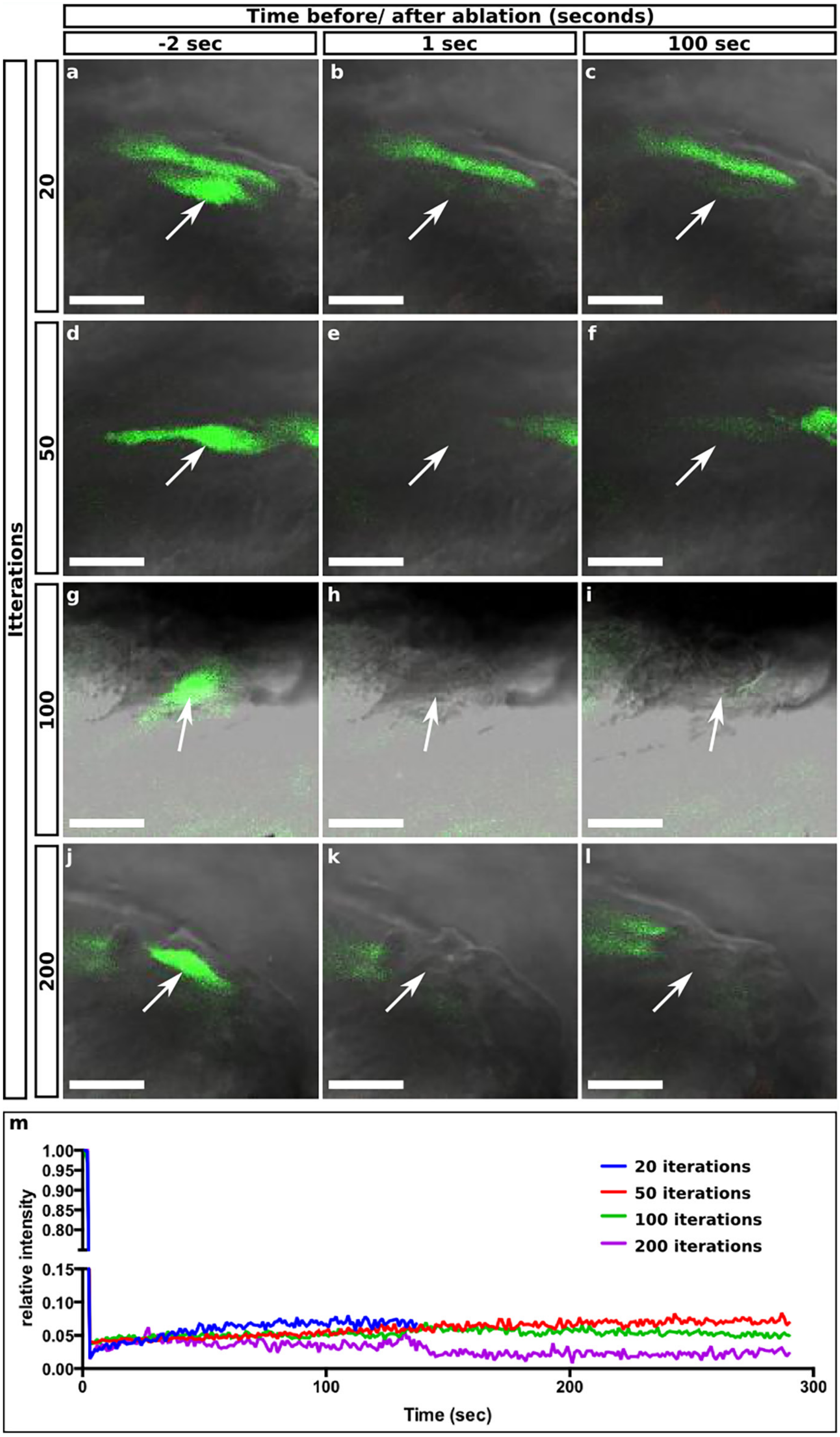
Extended Data Figure 4 | Validation of homotopic somite transplantations. **a–d**, Dorsal view of a somite 4 transplantation within a 13-somite stage embryo that has a Tg(*ubi*:zebrabow) (red) somite homotopically transplanted into a host embryo transgenic for Tg(*actc1b*:GFP), which marks the myotome in green. **e–h**, Dorsal view of the embryo in **a–d** at 24 hpf. **i–l**, Lateral view of the embryo in **a–d** at 2 dpf. **m**, Table of somite transplantations. Results from homotopic transplants of somites 3–6. **n–p**, Dorsal view of Tg(*actc1b*:GFP); Tg(*pax3a*:GFP) after somite 4 has been removed at the 15-somite stage. Solid lines indicate somite borders, numbers indicate somite identity. **q–r**, Lateral view of Tg(*actc1b*:GFP); Tg(*pax3a*:GFP) at 36 hpf. GFP-positive cells are present in the AER of both the control (**q**) and the experimental side (**r**) of the same embryo ($n = 5$). Scale bars, 100 μm (**a–i**), 50 μm (**n–p**) and 20 μm (**q**, **r**).

a

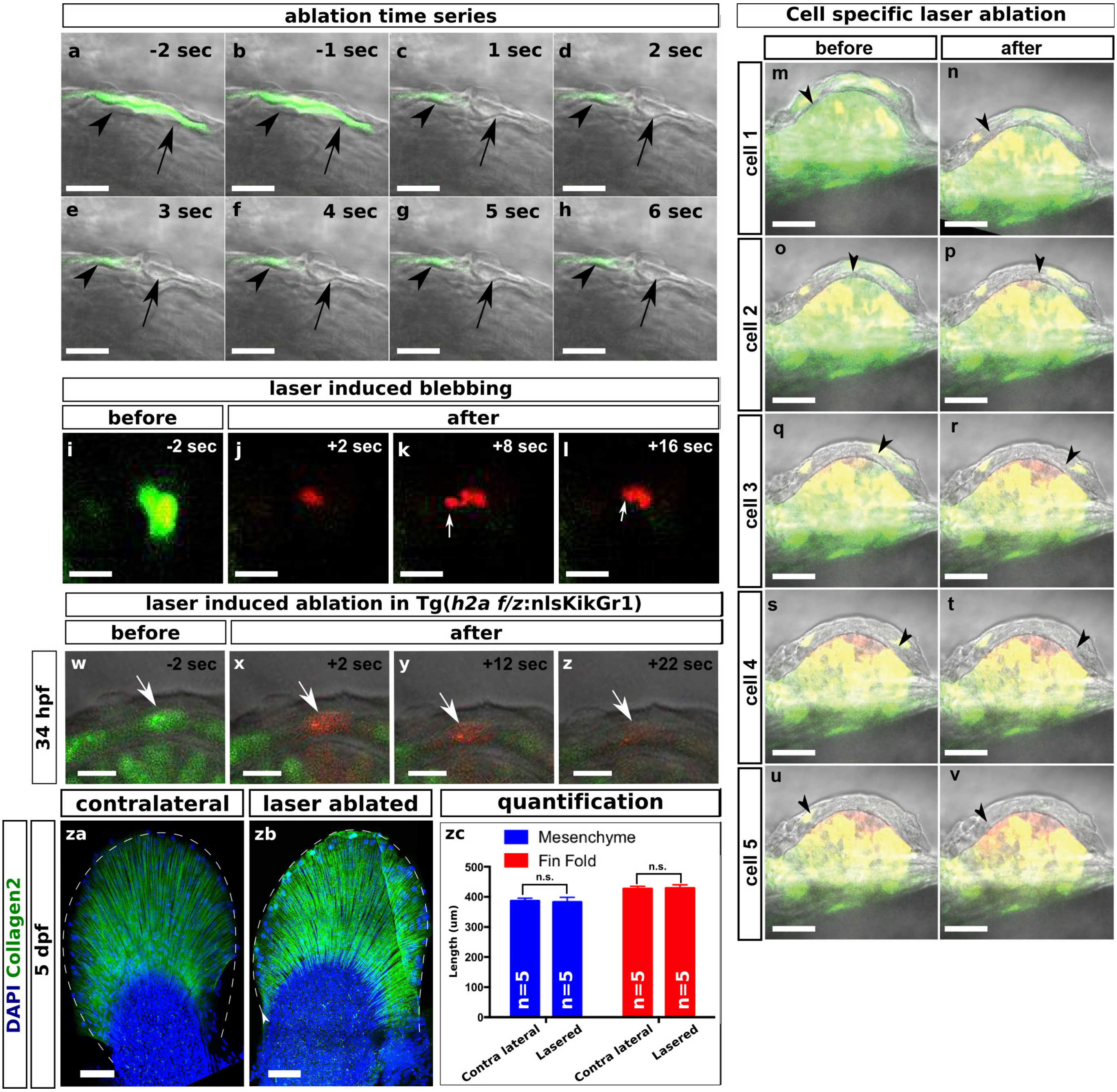
	0 mM Mtz (n=25)	20 mM Mtz (n=25)	40 mM Mtz (n=25)
Survival Rate	100%	92%	84%
Abnormal trunk muscle morphology	0%	4%	84%
NfsB-mCherry in AF	100%	8%	0%



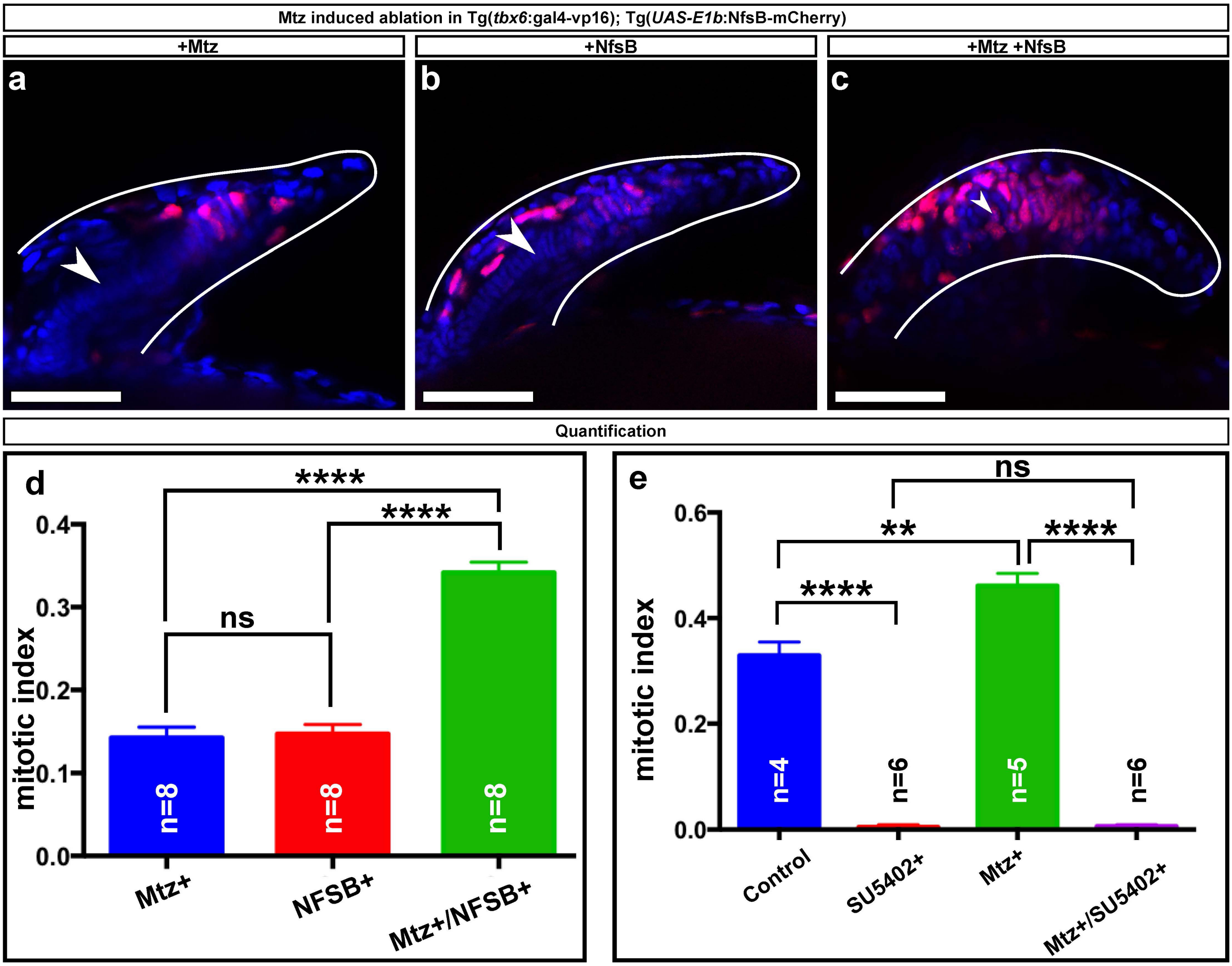
Extended Data Figure 5 | Optimization of MTZ-induced ablations. **a**–**d**, Tg(*tbx6*:gal4-vp16); Tg(*UAS-E1b*:Eco.NfsB-mCherry) embryos were treated with 0, 20 and 40 mM MTZ from 30–48 hpf and analysed at 48 hpf. **a**, Initial analyses assessed muscle transgene expression patterns, embryonic survival and overall phenotype severity as well as the efficiency of ablation of mCherry-expressing cells within the AF. **b**–**d**, Antibody stain for mCherry at 48 hpf reveals disrupted morphology of mesoderm-derived cells in the AF. Medio-lateral view of pectoral fins at 48 hpf treated with 0 mM MTZ ($n=8$) (**b**), 20 mM MTZ ($n=7$) (**c**), and 40 mM MTZ ($n=11$) (**d**). Arrows indicate dead or dying cells based on cell morphology and fragmentation, dashed lines indicate the border between the AF and underlying mesenchyme. Scale bars, 50 μm.



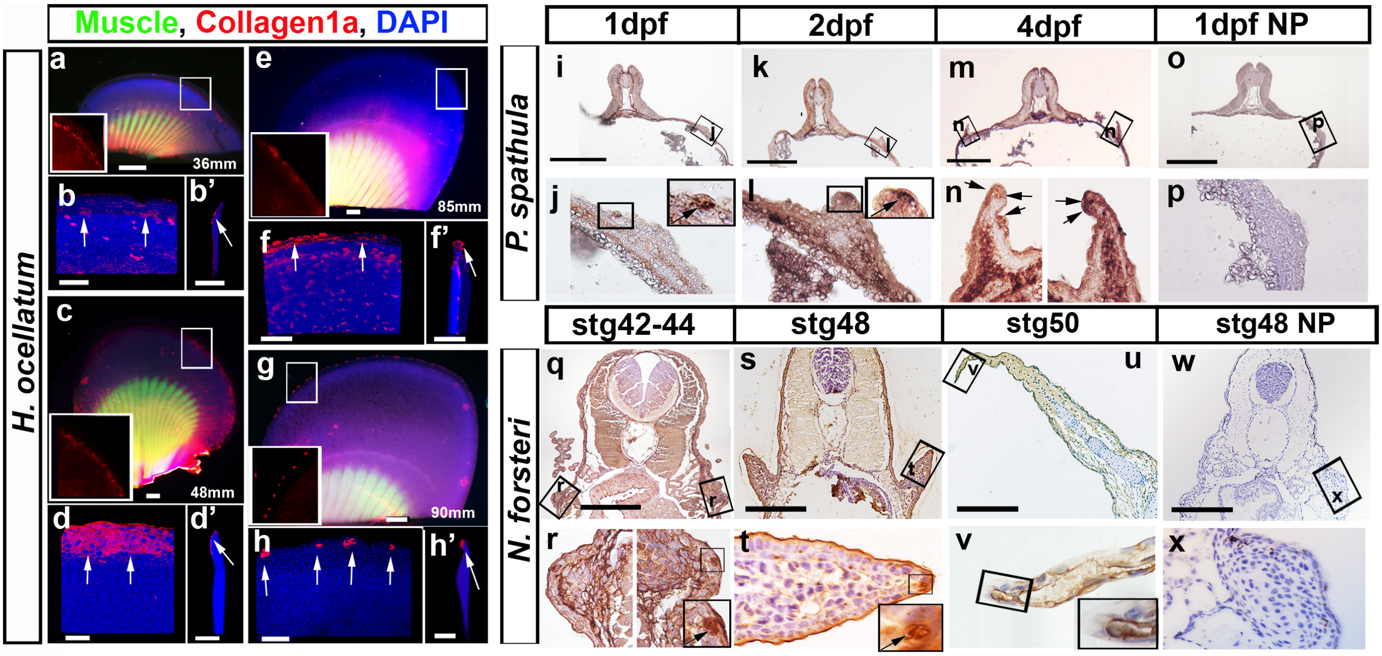
Extended Data Figure 6 | Optimization of the laser-ablation strategy. **a**–**l**, Assessment of fluorescent recovery after laser ablation of Kaede-positive cells in the AER of double-transgenic Tg(*tbx6*:gal4-vp16); Tg(*UAS*:Kaede) embryos by scanning the region of interest (ROI) for 20 (**a**–**c**), 50 (**d**–**f**), 100 (**g**–**i**) and 200 (**j**–**l**) laser pulse iterations. **m**, Fluorescent intensity of the ROI before (scaled to 1, the initial fluorescence of the targeted cells) and after laser-mediated cell ablation. A very small amount of fluorescent recovery can be observed at 20 (**c**, **m**), 50 (**f**, **m**) and 100 (**i**, **m**) scan iterations, while 200 scan iterations results in a consistent and permanent loss of Kaede fluorescence (**l**, **m**). Scale bars, 10 μm.



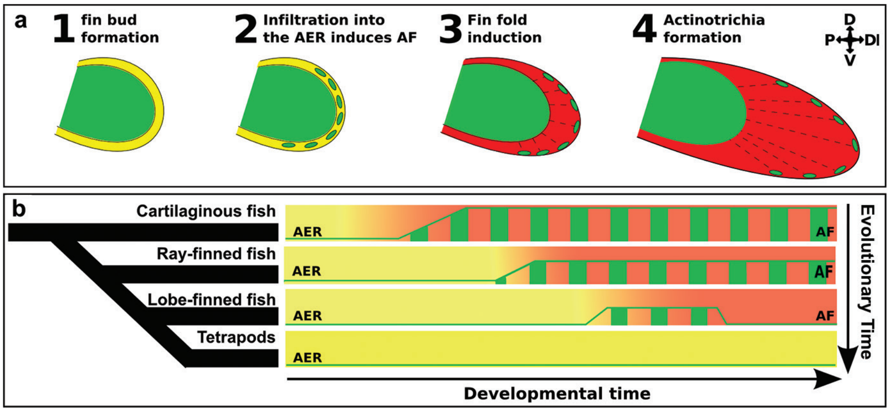
Extended Data Figure 7 | Validation of the laser-ablation strategy. **a**–**h**, Morphological disruption after laser ablation of Kaede-positive cells in the AER of Tg(*tbx6*:gal4-vp16); Tg(*UAS*:Kaede) double-transgenic embryos. Arrows mark the ablated cell while the adjacent cell, indicated by arrowheads, is not affected. **i**–**l**, Membrane blebbing, as visualized by photoconverted Kaede (red), can be observed in the AER of Tg(*tbx6*:gal4-vp16); Tg(*UAS*:Kaede) double-transgenic embryos. Before ablation the cell is clearly labelled by non-photoconverted Kaede (green) (**i**), whereas after ablation, blebbing is observed at selected time points (arrows) (**j**–**l**). **m**–**v**, Proof-of-concept experiment in which each Kaede-positive cell in the AER is sequentially ablated from Tg(*tbx6*:gal4-vp16); Tg(*UAS*:Kaede) double-transgenic embryos. Arrows mark ablated cells. Ablation is cell-specific, and does not occur in non-targeted Kaede-positive cells in the AER or the underlying mesenchyme. **w**–**zc**, Random ablation of AER cells before AFIC invasion does not affect fin formation. Using Tg(*h2a f/z*:nlsKikGR1) transgenic embryos, in which all nuclei are labelled by the photoconvertible protein Kikume, it is possible to ablate cells at 32 hpf within the AER before AFIC invasion. Arrows indicate targeted cells within the AER at 2 s before ablation (**w**) and after ablation at 2 s (**x**), 12 s (**y**) and 22 s (**z**). **za**–**zc**, In each experimental condition, 6 cells were ablated per AER. No differences in actinotrichia formation or overall morphology can be detected when comparing contralateral control (**za**) with laser-ablated (**zb**) conditions. Detailed quantification of fin fold and mesenchyme length does not reveal any significant differences between contralateral control and laser-ablated conditions (**zc**). *P* values are by unpaired Student's *t*-test. Scale bars, 10 μm (**a**–**l**), 50 μm (**m**–**v**), 10 μm (**w**–**z**) and 100 μm (**za**–**zc**). Error bars in **zc** represent s.e.m.



Extended Data Figure 8 | Ablation of somite-derived cells from the AER results in increased proliferation of the underlying mesenchyme. **a–d**, Ablation of somite-derived cells in the AER in embryos transgenic for Tg(*tbx6*:gal4-vp16) and Tg(*UAS*:Kaede) results in an increased mitotic index of the underlying mesenchyme as measured by EdU pulse labelling. All samples were analysed at 48 hpf. The endoskeletal disk is undergoing condensation in controls (arrowheads) (**a**, **b**), while this is not evident in MTZ-treated animals (arrow) (**c**). **d**, Quantification of EdU staining observed in the mesenchyme of pectoral fins after MTZ-induced ablation of somite-derived cells from the AER. **e**, Inhibition of Fgf signalling reduced the mitotic index in pectoral fins that have undergone MTZ-mediated ablation of somite-derived cells in the AER. The mitotic index of control fins is significantly lower compared to MTZ-treated pectoral fins. However, after treatment with the Fgf inhibitor SU5402, both MTZ-treated and untreated pectoral fins are reduced to a similar mitotic index. Scale bars, 40 μm. *P* values are by ANOVA with Tukey's post-hoc analysis. Error bars in **d** and **e** represent s.e.m.



Extended Data Figure 9 | Progressive reduction of collagen secreting cells from the AER of different species in the gnathostome phylogeny. **a**–**h**, Embryos of the epaulette shark *H. ocellatum* stained for Col1a1a (red), skeletal muscle (green) and nuclei (blue). **a**, **b**, At 36 mm (approximately *S. canicula* stage 26/27; ref. 32), Col1a1-expressing cells are present at low numbers within the AER initially, with cells evident traversing the AER basement membrane from the fin mesenchyme (arrows in **b**, **b'**). The density of Col1a1a-expressing cells within the AER increases notably with the extension of the fin, and the cells are present in large numbers in 48 mm embryos (**c**, **d**), (approximately *S. canicula* stage 30/31; ref. 28). Col1a1a-expressing cells are maintained at the edge of the fin fold at advanced stages of fin development (85 mm, **e**, **f**) and cluster into discrete foci at the edge of the fin fold by 90 mm (**g**, **h**). **a**, **c**, **e**, **g**, Dorsal views proximal to the top, inserts are high-magnification views of the boxed regions showing only Col1a1a expression in red in each fin. **b**, **d**, **f**, **h**, Maximum projection confocal renders of the distal edge of the AER/AF region of each fin in **a**, **c**, **e** and **g**, respectively. **b'**, **d'**, **f'**, **h'**, Sagittal views of the rendered images in **b**, **d**, **f** and **h**, respectively, revealing the persistent localization of Col1a1a-expressing cells at the distal most edge of each fin (arrows). **i**–**p**, Embryos of the paddlefish *P. spathula*, sectioned in the transverse plane and stained for Col1a1a expression (brown, DAB) at 1 dpf, (**i**, **j**), 2 dpf (**k**, **l**) and 4 dpf (**m**, **n**). Colonization of the AER by Col1a1a-expressing cells follows a very similar pattern temporally and spatially to that evident in the zebrafish fin, but appear at a lower density than in the epaulette shark. **o**, **p**, Control immunohistochemistry on 1-dpf sections with no primary antibody. **j**, **l**, **n**, **p**, High-magnification views of the boxed regions in **i**, **k**, **m** and **o**, respectively. Arrows in **j**, **l** and **n** mark Col1a1a-expressing cells in the AER. **q**–**x**, *N. forsteri* embryos in transverse section and stained for Col1a1a and counterstained with H&E. Very few Col1a1a-expressing cells colonize the AER initially (**q**, **r**, stage 42–44), and they continue to be present in low numbers up to stage 48 (**s**, **t**) and are no longer evident at fin extension and AF formation stages (**u**, **v**, stage 50). **w**, **x**, Control immunohistochemistry on stage-48 sections with no primary antibody. **r**, **t**, **v**, **x**, High-magnification views of the boxed regions in **q**, **s**, **u** and **w**, respectively. Arrows in inset images in **r** and **t** mark the limited number of Col1a1a-expressing cells in the AER of lungfish embryos. Scale bars, 200 µm (**a**, **c**), 50 µm (**b**, **f**), 100 µm (**d**, **h**), 400 µm (**e**) and 500 µm (**g**, **i**, **k**, **m**, **o**, **q**, **s**, **u**, **w**).



Extended Data Figure 10 | Schematic representation of pectoral fin development and its relation to the fin-to-limb transition.
a, Induction of fin fold formation and subsequent actinotrichia formation are dependent on somitic mesoderm-derived apical fold inducing cells (AFICs) cells infiltrating the AER. After the induction of fin bud formation (1), a somite-derived lineage (green) will infiltrate the AER (yellow, 2). Subsequent fin fold (red) induction (3) is dependent on this process, as is the formation of actinotrichia (dashed lines, 3 and 4), which rely on the secretion of collagen directly from these cells (green, 3 and 4). **b**, Throughout evolution of the gnathostome lineage, the cells that regulate fin fold induction have been gradually lost as a necessary prerequisite for limb formation. Schematic representation of fin development including the transition from AER (yellow) to AF (red). The timing of somite-derived cells (green bars) infiltrating the AER coincides with the AER to AF transition. Relative numbers of somite-derived cells are represented by the height of the green bars. In cartilaginous fish, cells infiltrate the AER early and abundantly to remain in the fin fold throughout development. In ray-finned fish, infiltration into the AER is delayed and the number of cells infiltrating the AER is also reduced, but infiltration continues throughout development. The number of cells infiltrating the AER and their timing of residency is further reduced in lobe-finned fish, while finally in tetrapods there is an absence of somite-derived cells in the AER, and a lack of AF induction.

doi:10.1038/nature18635

Indifference to dissonance in native Amazonians reveals cultural variation in music perception

Josh H. McDermott[1], Alan F. Schultz[2], Eduardo A. Undurraga[3,4] & Ricardo A. Godoy[3]

Music is present in every culture, but the degree to which it is shaped by biology remains debated. One widely discussed phenomenon is that some combinations of notes are perceived by Westerners as pleasant, or consonant, whereas others are perceived as unpleasant, or dissonant[1]. The contrast between consonance and dissonance is central to Western music[2,3], and its origins have fascinated scholars since the ancient Greeks[4–10]. Aesthetic responses to consonance are commonly assumed by scientists to have biological roots[11–14], and thus to be universally present in humans[15,16]. Ethnomusicologists[17] and composers[8], in contrast, have argued that consonance is a creation of Western musical culture[6]. The issue has remained unresolved, partly because little is known about the extent of cross-cultural variation in consonance preferences[18]. Here we report experiments with the Tsimane'—a native Amazonian society with minimal exposure to Western culture—and comparison populations in Bolivia and the United States that varied in exposure to Western music. Participants rated the pleasantness of sounds. Despite exhibiting Western-like discrimination abilities and Western-like aesthetic responses to familiar sounds and acoustic roughness, the Tsimane' rated consonant and dissonant chords and vocal harmonies as equally pleasant. By contrast, Bolivian city- and town-dwellers exhibited significant preferences for consonance, albeit to a lesser degree than US residents. The results indicate that consonance preferences can be absent in cultures sufficiently isolated from Western music, and are thus unlikely to reflect innate biases or exposure to harmonic natural sounds. The observed variation in preferences is presumably determined by exposure to musical harmony, suggesting that culture has a dominant role in shaping aesthetic responses to music.

We conducted two studies to measure consonance preferences in populations with varying exposure to Western music. In Study 1, we measured preferences for sounds in residents of the United States and compared them to three populations in Bolivia: (1) residents of the capital city (La Paz); (2) residents of a rural town (San Borja); and (3) members of a native society of horticulturalist-foragers (the Tsimane') in a remote village (Santa Maria) in the Amazon rainforest. City- and town-dwellers were fluent in Spanish and generally had televisions and radios. By contrast, the Tsimane' were mostly monolingual in their own language, lacked televisions, and had limited access to music via radio[19]. The Tsimane' village lacked electricity and tap water, was inaccessible by road, and could be reached only by canoe (Fig. 1a). Their contact with Western culture was mainly limited to occasional trips to nearby towns. To compare differences between cultures to intra-cultural variation within Westerners, we tested two groups from the United States—one with at least two years of experience playing a musical instrument, and one with at most one year of experience.

The Tsimane' were of particular interest because harmony, polyphony, and group performances are by all accounts absent from their music. This conclusion was suggested by previous recordings and documentation[20] and was substantiated by interviews with Tsimane' musicians (see Methods for a description of Tsimane' music, and Supplementary Audio 1 and 2 for examples of Tsimane' songs). The apparent absence of harmony raised the question of whether they would exhibit aesthetic responses to consonance and dissonance despite not having prior exposure to them.

Participants were presented with sounds over headphones (Fig. 1b) and rated the pleasantness of sounds on a four-point scale. We observed pronounced differences between groups in the ratings of consonant and dissonant chords (Fig. 2a, b). US residents showed strong consonance preferences, as expected from previous studies[21], with musicians showing stronger preferences than non-musicians (synthetic chords: $F(46,1) = 7.44$, $P = 0.009$; sung chords: $F(46,1) = 12.97$, $P = 0.001$). Moreover, Bolivian city- and town-dwellers displayed more modest but statistically significant preferences for consonance ($P < 0.01$ in all cases, *t*-tests). By contrast, the Tsimane' rated consonant and dissonant chords as equally pleasant. This finding held for chords composed of synthetic tones as well as chords composed of recorded notes sung by a vocalist, producing an interaction in both cases between stimulus class and participant group (synthetic: $F(124,4) = 48.43$, $P < 10^{-23}$; sung: $F(126,4) = 26.79$, $P < 10^{-15}$). Tsimane' ratings varied somewhat across chords (synthetic: $F(9,270) = 1.96$, $P = 0.04$; sung: $F(9,288) = 1.93$, $P = 0.05$), with some tendency for larger pitch intervals to be preferred over smaller ones (Extended Data Fig. 1), but without consistent differences between consonant and dissonant chords.

Similar results were obtained when participants rated single tones whose frequencies were manipulated to be harmonic (multiples of a common fundamental frequency) or inharmonic, a distinction believed to differentiate consonant from dissonant chords[4,12,13,21,22] (Fig. 2c). US residents showed a preference for harmonic over inharmonic tones, but the effect was stronger in musicians than in non-musicians ($F(46,2) = 16.12$, $P = 0.0002$), was attenuated in Bolivian city-dwellers, and was undetectable in town-dwellers and the Tsimane', again producing an interaction between stimulus and participant group ($F(157,4) = 22.49$, $P < 10^{-13}$).

To rule out the possibility that participant groups might have varied in their understanding of task instructions, or in their ability to assign pleasantness ratings to sounds, we asked them to rate recordings of emotional vocalizations[22] (Fig. 2d). In this case all groups showed a preference for laughter over gasps ($F(157,1) = 200.48$, $P < 1 \times 10^{-29}$), with no interaction with group ($F(157,4) = 1.09$, $P = 0.36$). Thus, for familiar sounds expected to have a positive or negative association, the Tsimane' exhibited a preference in the expected direction, comparable to that of the other groups.

To test whether the Tsimane' would generally fail to exhibit preferences for unfamiliar sounds, we obtained ratings for synthetic tones that varied in roughness, another quality known to affect aesthetic responses in Westerners[5,7,21–25] (Fig. 2e). Roughness was manipulated by presenting pairs of frequencies to the same or different ears. In

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Department of Anthropology, Baylor University, Waco, Texas 76798, USA. [3]Heller School for Social Policy and Management, Brandeis University, Waltham, Massachusetts 02453, USA. [4]Center for Intercultural and Indigenous Research, Pontificia Universidad Católica de Chile, Santiago, Región Metropolitana 7820436, Chile.

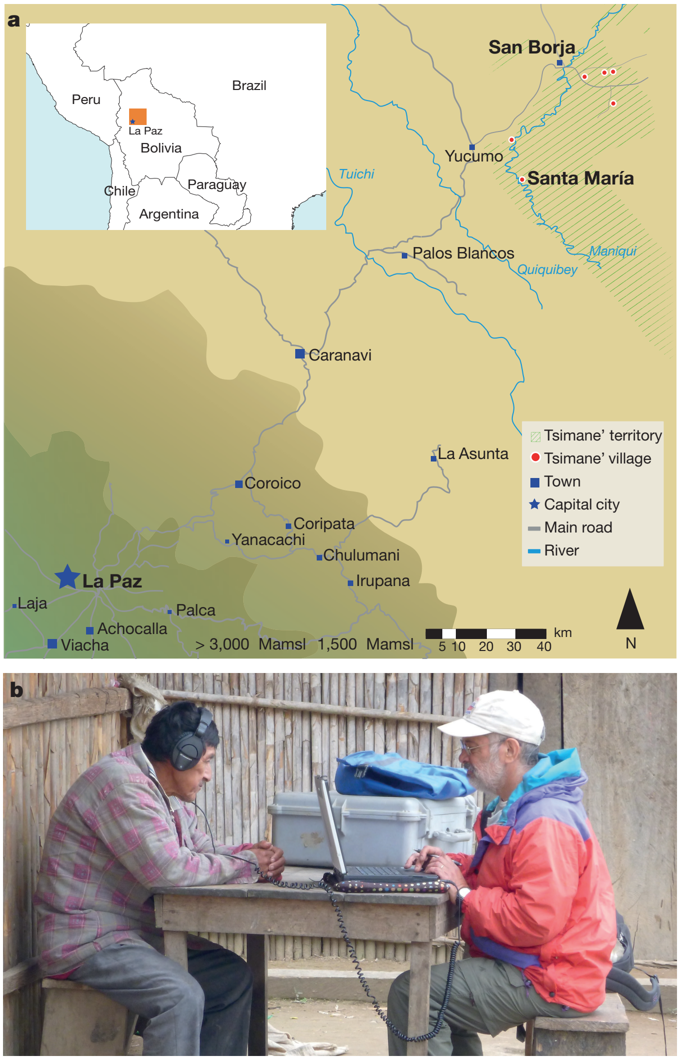


Figure 1 | Location and setup of experiments. a, Map of the region from which Bolivian participants were drawn. Participants resided in the Bolivian capital city La Paz, the rural town of San Borja, the Tsimane’ village of Santa Maria (Study 1), or Tsimane’ villages around San Borja (Study 2; not labelled with names to minimize clutter). Town symbols are approximately proportional in size to town population. Colour of territory denotes elevation. **b,** Sounds were presented over closed headphones via laptop (charged with a gasoline generator when needed). For all but the discrimination experiment (Fig. 4), participants provided a pleasantness rating (with a four-point scale) following each sound.

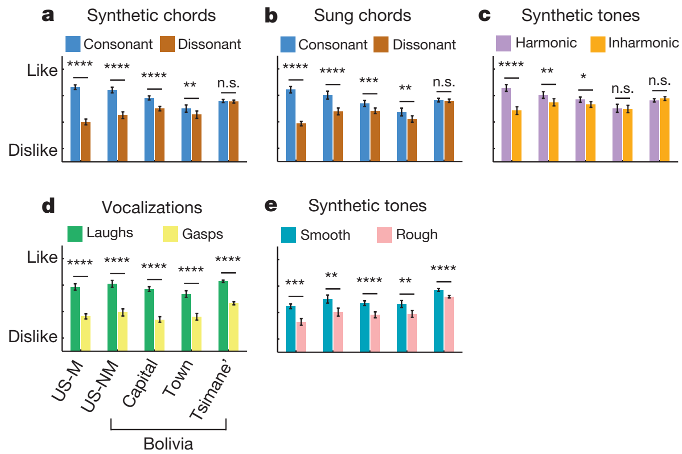


Figure 2 | Results of Study 1. a–e, Average pleasantness ratings of sounds from five experiments across five populations: 23 US musicians (US-M), 25 US non-musicians (US-NM), 24 Bolivian city-dwellers (Capital), 26 Bolivian town-dwellers, and 64 Tsimane’. Each experiment featured sounds from two classes expected to differ in pleasantness for US listeners. Chord notes were either synthetic (resembling a piano), or recorded from a trained singer. Chords were conventionally consonant (major third, perfect fourth, perfect fifth, and major triad) or conventionally dissonant (minor second, major second, tritone, major seventh, and augmented triad). Vocalizations were recordings of human laughs and gasps. Synthetic tones varied in harmonicity or roughness. Asterisks denote statistical significance: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; n.s., not significant (two-tailed t-tests, uncorrected for multiple comparisons because they were conducted post-hoc, following analysis of variances (ANOVAs) to test for main effects and interactions). Data are mean and s.e.m.

contrast to the results for chords, and for harmonic and inharmonic tones, all groups exhibited a significant preference for smooth over rough tones ($F(157,1) = 85.51$, $P < 10^{-15}$), with no interaction with group ($F(157,4) = 2.04$, $P = 0.09$), even though for US participants this preference was the smallest of those measured.

We replicated and extended these results in Study 2, testing a different group of Tsimane’ listeners and a comparison group of musically experienced listeners in the United States. We first tested whether the absence of a consonance preference would extend to harmony in realistic musical material. We recorded Tsimane’ vocalists singing Tsimane’ song phrases several times, and then pitch-shifted and superimposed the phrases to create harmonies in conventionally consonant or dissonant intervals (see Supplementary Audio 3–10). We obtained pleasantness ratings for these harmonies from US and Tsimane’ listeners as in Study 1. For replication purposes, we also conducted analogous experiments with sung and synthetic two-note intervals, and with sung triads.

The results with chords replicated those of Study 1 (Fig. 3a–c): in every case, US listeners preferred consonance to dissonance, while Tsimane’ listeners did not, producing interactions between stimulus and participant group (synthetic intervals: $F(72,1) = 42.4$, $P < 10^{-8}$; sung intervals: $F(95,1) = 39.4$, $P < 10^{-7}$; sung triads: $F(95,1) = 17.9$, $P < 10^{-4}$). Ratings of individual chords by Tsimane’ listeners again varied across both synthetic and sung intervals (Extended Data Fig. 2), but were primarily explained by interval size, with higher ratings for larger intervals, unlike US listeners (significant correlations between rating and interval size for Tsimane; synthetic: $r = 0.94$, $P < 10^{-5}$; sung: $r = 0.81$, $P = 0.001$; but not for US listeners; synthetic: $r = 0.07$, $P = 0.82$; sung: $r = 0.23$, $P = 0.48$).

Notably, similar results were obtained with harmonies generated from Tsimane’ songs (Fig. 3d). Even though the music was foreign to US participants, they reliably judged consonant renditions as more pleasant than dissonant ($t(46) = 6.2$, $P < 10^{-6}$), whereas the Tsimane’ did not ($t(49) = 1.2$, $P = 0.22$; stimulus × group interaction: $F(95,1) = 30.2$, $P < 10^{-6}$). Moreover, Tsimane’ listeners reliably preferred some of the song excerpts used to generate harmonies over others ($\chi^2(25) = 49.01$, $P = 0.003$; Extended Data Fig. 3). The materials thus elicited consistent aesthetic responses in the Tsimane’, but these were not driven by consonance and dissonance. As in Study 1, when presented with recorded vocalizations (Fig. 3e), both Tsimane’ and US listeners showed preferences for laughter over gasps ($F(95,1) = 129.4$, $P < 10^{-18}$; no interaction with participant group: $F(95,1) = 1.8$, $P = 0.18$), indicating that Tsimane’ listeners could readily perform the task. These results again suggest that the preference for consonance is absent in the Tsimane’.

To explore the effects of harmonicity and roughness found in Study 1, we measured pleasantness ratings for pairs of pure tones (single frequencies) separated by intervals from the chromatic scale (0–8 semitones)[21] (Fig. 3f). This range includes some consonant intervals, for which the tone frequencies approximate harmonics of a common fundamental (and are thus related by simple integer ratios), and some dissonant intervals, for which the tone frequencies are inharmonic. Headphones were used to present the two tones to both ears (diotic presentation) or to separate ears (dichotic presentation), as in the smooth/rough tone experiment from Study 1. Diotic presentation

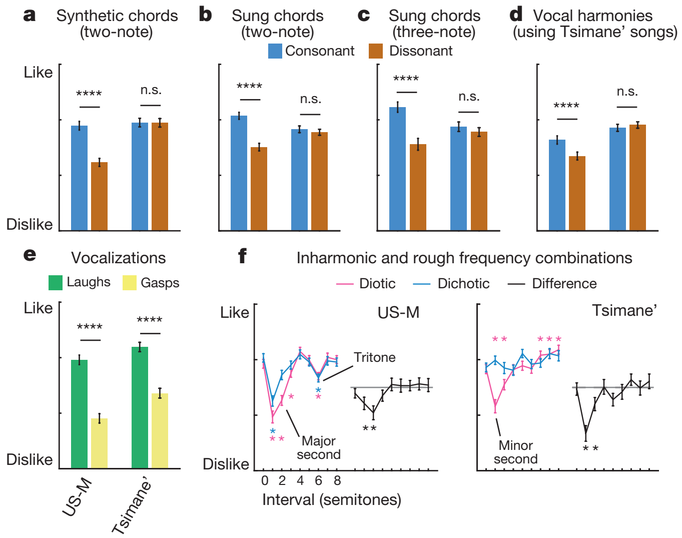


Figure 3 | Results of preference experiments from Study 2. a–d, Average pleasantness ratings of consonant and dissonant musical stimuli by 47 US musicians (US-M) and 50 Tsimane'. Two-note chords were intervals from the chromatic scale. Triads were major or augmented. Vocal harmonies (intervals from the chromatic scale) were generated from recordings of Tsimane' vocalists. **e**, Average pleasantness ratings of human laughs and gasps. Asterisks denote statistical significance: ****$P < 0.0001$ (two-tailed t-tests) (**a**–**e**). **f**, Average pleasantness ratings of concurrent pure tones (single frequencies)[21] presented diotically (both tones to both ears) or dichotically (each tone to a different ear). Blue and pink asterisks indicate ratings significantly different from that of the unison ($P < 0.05$; Wilcoxon signed rank tests, two-tailed, uncorrected for multiple comparisons). Black curves plot the difference between ratings of dichotic and diotic stimuli. Black asterisks denote dichotic–diotic differences significantly different from zero ($P < 0.05$; Wilcoxon signed rank tests, two-tailed, uncorrected). Data are mean and s.e.m.

produces roughness (via beating) when the tones are close in frequency, whereas dichotic presentation minimizes beating while preserving the musical interval (ratio) between frequencies. The stimuli thus dissociated roughness and harmonicity.

Among US residents, pleasantness ratings dipped at dissonant intervals (the minor second, major second, and tritone) for both diotic and dichotic presentation. The variation in pleasantness for dichotic intervals ($\chi^2(7) = 62.7$, $P < 10^{-10}$) implicates factors other than roughness in listeners' aesthetic judgments, potentially that of harmonic frequency relations. However, for small intervals (<3 semitones), diotic ratings were nonetheless lower than dichotic ratings, evident in the difference curve, which varied significantly with interval ($\chi^2(7) = 20.6$, $P = 0.004$; significant differences from zero only for 1 and 2 semitones, $P < 0.05$). This effect is the signature of an aversion to roughness, which is audible only for small intervals (for which frequencies interact via cochlear filtering)[5,7,21].

Tsimane' ratings also suggested an aversion to roughness: diotic intervals were again rated lower than dichotic only for small intervals, producing significant variation in the diotic-dichotic difference ($\chi^2(7) = 29.25$, $P = 0.0001$; significant differences again only for 1 and 2 semitones, $P < 0.05$). By contrast, the influence of harmonicity was absent. Ratings of dichotic intervals did not vary with interval ($\chi^2(7) = 9.52$, $P = 0.22$), and the dip at the tritone was not present for either presentation mode. These results provide further evidence that an aesthetic aversion to roughness is present in the Tsimane', but that the aesthetic response to consonance (and simple integer frequency ratios) is not. This dissociation substantiates the conclusion that the Western notion of dissonance is distinct from acoustic roughness, but closely related to inharmonicity[21,22].

The Tsimane' indifference to consonance and harmonicity raises the question of whether they can hear the underlying stimulus distinction.

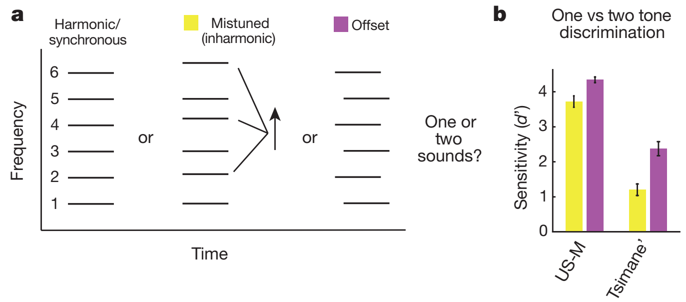


Figure 4 | Results of discrimination experiment from Study 2. **a**, Participants (47 US musicians and 49 Tsimane') heard six harmonics and judged whether they heard one or two sounds. On half of the trials the even harmonics were mistuned by 12% or presented 150 ms earlier in time (harmonics were 750 ms in duration). **b**, Ability to discriminate one vs. two sounds by harmonicity or onset differences. Chance performance corresponds to $d' = 0$. Data are mean and s.e.m.

To test their ability to discern harmonic from inharmonic frequency relations, we conducted a discrimination experiment in which listeners judged whether a stimulus contained one or two sounds (Fig. 4a). Listeners heard the first six harmonics of a synthetic tone. On half of the trials, the even harmonics were mistuned or presented earlier in time. Both manipulations used an established sound segregation cue to cause the even and odd harmonics to be heard as distinct sounds[26,27]. In particular, the mistuning disrupted harmonicity, causing the even harmonics to have a different fundamental frequency than the odd harmonics, producing the perception of a dissonant interval in Western listeners.

US residents reliably identified the intended number of tones in both conditions, as expected (Fig. 4b, left: $t(46) = 22.9$, $P < 10^{-26}$; $t(46) = 54.4$, $P < 10^{-42}$). Tsimane' listeners were worse overall at the task, but were well above chance in both cases (Fig. 4b, right: $t(48) = 7.3$, $P < 10^{-8}$; $t(48) = 11.8$, $P < 10^{-15}$), indicating sensitivity to harmonicity. Analysis of subsets of Tsimane' participants revealed that even those who performed best on this task (mean sensitivity index (d') of 2.0 for the mistuned condition) remained indifferent to dissonance, dissociating perceptual sensitivity and preference (Extended Data Fig. 4).

Taken together, the results suggest that the Tsimane' are able to make pleasantness judgments about sounds. Moreover, for familiar sounds associated with a valence (laughs and gasps), and for synthetic sounds varying in roughness, their judgments resemble those of Westerners. In addition, they are able to hear the acoustic distinctions associated with consonance and dissonance, discriminating harmonic from inharmonic frequencies. The absence of a measurable Tsimane' preference for consonance over dissonance thus appears to specifically reflect differences in their aesthetic response to this contrast.

The cross-cultural variation we observed suggests that consonance preferences are unlikely to be innate, and that they are not driven by exposure to harmonic natural sounds such as vocalizations[14]. Instead, consonance preferences seem to depend on exposure to particular types of music, presumably those that feature consonant harmony. Incidental exposure to Western music could explain the modest preferences evident in Bolivian city- and town-dwellers, along with those previously reported in human infants[28–30]. The lack of harmony in Tsimane' music leaves open the question of what preferences might be observed in individuals exposed to conventionally dissonant intervals, as are prevalent in some non-Western music[17]. But Western music is presumably not unique in fostering preferences for some tone combinations over others.

The belief that consonance is biologically determined has been fuelled by observations that consonance and dissonance differ in acoustic properties such as harmonicity and roughness[4,5,7,9,12,21,22], and as such are not arbitrary categories. However, we found preferences for harmonic tones, for simple integer frequency ratios, and for consonant chords/harmonies to all co-vary with presumptive exposure to

Western culture. Thus, rather than being an inevitable consequence of auditory system biology, it seems that the preferences exhibited by Western listeners for harmonic frequencies arise from exposure to Western music because the musical structures prevalent therein tend to produce harmonic frequencies[21,22]. And although an aversion to roughness is apparently present cross-culturally, it seems to be unrelated to consonance and dissonance, perhaps because musical sounds are in practice not very rough[21,22]. The role of biology in consonance could therefore be limited to constraining what is discriminable, potentially rendering some aesthetic contrasts more readily acquired than others.

The roles of culture and biology in consonance and dissonance have remained unresolved in part because cross-cultural experimental data are rare. Studies of culturally isolated populations are increasingly difficult to conduct due to the diffusion of Western culture around the world, but our results underscore their importance for revealing the diversity of human musical behaviour.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 November 2014; accepted 10 June 2016.
Published online 13 July 2016.

1. Parncutt, R. & Hair, G. Consonance and dissonance in theory and psychology: Disentangling dissonant dichotomies. *J. Interdiscipl. Music Stud*. **5**, 119–166 (2011).
2. Huron, D. Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Percept.* **11,** 289–305 (1994).
3. Bigand, E., Parncutt, R. & Lerdahl, F. Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* **58,** 124–141 (1996).
4. Stumpf, C. *Tonpsychologie* (Verlag S. Hirzel, 1890).
5. von Helmholtz, H. *Die Lehre von den Tonempfindungen als physiologische Grundlage fur die Theorie der Musik* (F. Vieweg und Sohn, 1863).
6. Lundin, R. W. Toward a cultural theory of consonance. *J. Psychol.* **23,** 45–49 (1947).
7. Plomp, R. & Levelt, W. J. M. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38,** 548–560 (1965).
8. Cazden, N. The definition of consonance and dissonance. *Int. Rev. Aesthet. Soc.* **11,** 123–168 (1980).
9. Sethares, W. A. *Tuning, Timbre, Spectrum, Scale* (Springer, 1999).
10. Tenney, J. *A History of 'Consonance' and 'Dissonance'* (Excelsior Music Publishing Company, 1988).
11. Fishman, Y. I. *et al.* Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *J. Neurophysiol.* **86,** 2761–2788 (2001).
12. Tramo, M. J., Cariani, P. A., Delgutte, B. & Braida, L. D. Neurobiological foundations for the theory of harmony in western tonal music. *Ann. NY Acad. Sci.* **930,** 92–116 (2001).
13. Bidelman, G. M. & Heinz, M. G. Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. *J. Acoust. Soc. Am.* **130,** 1488–1502 (2011).
14. Bowling, D. L. & Purves, D. A biological rationale for musical consonance. *Proc. Natl Acad. Sci. USA* **112,** 11155–11160 (2015).
15. Butler, J. W. & Daston, P. G. Musical consonance as musical preference: a cross-cultural study. *J. Gen. Psychol.* **79,** 129–142 (1968).
16. Fritz, T. *et al.* Universal recognition of three basic emotions in music. *Curr. Biol.* **19,** 573–576 (2009).
17. Brown, S. & Jordania, J. Universals in the world's musics. *Psychol. Music* **41,** 229–248 (2013).
18. Maher, T. F. "Need for resolution" ratings for harmonic musical intervals: A comparison between Indians and Canadians. *J. Cross Cult. Psychol.* **7,** 259–276 (1976).
19. Godoy, R. *et al.* Moving beyond a snapshot to understand changes in the well-being of native Amazonians. *Curr. Anthropol.* **50,** 563–573 (2009).
20. Riester, J. *Canción y Producción en la Vida de un Pueblo Indígena: los Chimane del Oriente Boliviano* (Los Amigos del Libro, 1978).
21. McDermott, J. H., Lehr, A. J. & Oxenham, A. J. Individual differences reveal the basis of consonance. *Curr. Biol.* **20,** 1035–1041 (2010).
22. Cousineau, M., McDermott, J. H. & Peretz, I. The basis of musical consonance as revealed by congenital amusia. *Proc. Natl Acad. Sci. USA* **109,** 19858–19863 (2012).
23. Terhardt, E. On the perception of periodic sound fluctuations (roughness). *Acustica* **30,** 201–213 (1974).
24. Kumar, S., Forster, H. M., Bailey, P. & Griffiths, T. D. Mapping unpleasantness of sounds to their auditory representation. *J. Acoust. Soc. Am.* **124,** 3810–3817 (2008).
25. Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L. & Poeppel, D. Human screams occupy a privileged niche in the communication soundscape. *Curr. Biol.* **25,** 2051–2056 (2015).
26. Moore, B. C., Glasberg, B. R. & Peters, R. W. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* **80,** 479–483 (1986).
27. Darwin, C. J. Perceiving vowels in the presence of another sound: constraints on formant perception. *J. Acoust. Soc. Am.* **76,** 1636–1647 (1984).
28. Zentner, M. R. & Kagan, J. Perception of music by infants. *Nature* **383,** 29 (1996).
29. Trainor, L. J., Tsang, C. D. & Cheung, V. H. W. Preference for sensory consonance in 2- and 4-month-old infants. *Music Percept.* **20,** 187–194 (2002).
30. Plantinga, J. & Trehub, S. E. Revisiting the innate preference for consonance. *J. Exp. Psychol. Hum. Percept. Perform.* **40,** 40–49 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors gratefully acknowledge the support of the National Science Foundation to R.A.G., a McDonnell Scholar Award to J.H.M., the TAPS Bolivia Study Team (particularly T. Huanca), C. García for assistance with Fig. 1a, E. Gibson for logistical help, N. Jacoby and M. Salinas for assistance recording and interviewing Tsimane' musicians, S. Popham and L. Chen for data collection, and D. Boebinger and K. Woods for comments on the manuscript.

Author Contributions J.H.M., A.F.S., E.A.U. and R.A.G. designed the experiments, collected the data, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.M. (jhm@mit.edu).

Reviewer Information *Nature* thanks S. Trehub and the anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Participants (Study 1). *US musicians.* The first group of US residents ($n = 23$, 19 female; mean age = 20.2 years, s.d. = 2.4 years) was selected to have experience playing a musical instrument (mean number of years spent playing an instrument = 7.7 years, s.d. = 4.9 years, range = 2–18 years). None were professional musicians, but all reported having taken music lessons or classes for much of the time that they were musically active (mean number of years during which lessons were taken = 6.8 years, s.d. = 4.3 years). All were residents of New York City. Two participants had immigrated to the United States as young children (at the ages of 2 and 4 years old); the rest were born and raised in the United States.

US non-musicians. The second group of US residents ($n = 25$, 12 female; mean age = 37.9 years, s.d. = 15.1 years) was selected to have minimal experience playing or studying music (mean number of years spent playing an instrument or singing = 0.2 years, s.d. = 0.3 years, range = 0–1 years). 17 out of 25 (68%) had never played an instrument or sung in an organized setting, and most had never taken music lessons or classes (mean number of years during which lessons were taken = 0.1 years, s.d. = 0.3 years). All were residents of the Boston metropolitan area and had been born and raised in the United States. No attempt was made to achieve demographic diversity among US participants because numerous similar experiments had previously been conducted in a variety of populations from Western societies[21,22,31–33], and the effects are known to generalize across age, gender, and geographical region of North America. However, the non-musician US sample was similar in demographic composition and musical experience to each of the Bolivian populations tested.

La Paz (capital). Participants ($n = 24$, 14 female; mean age = 29.9 years, s.d. = 9.3 years) had been born and raised in La Paz, the capital of Bolivia. Only three reported any experience playing music (across the group: mean number of years spent playing an instrument = 0.3 years, s.d. = 0.8 years), and only five reported experience singing (across the group: mean number of years spent singing = 0.5 years, s.d. = 1.1 years).

San Borja (town). Participants ($n = 26$, 15 female; mean age = 31.6 years, s.d. = 12.9 years) had resided for most of their life in San Borja, a small town in the department of Beni in the Amazon basin. San Borja could be reached by car during dry months of the year from other towns in Bolivia, but was accessible only by plane during much of the rainy season owing to land slides and flooding. Most of the townspeople tested (17 out of 26) reported having experience singing, but only 8 out of 26 had experience playing a musical instrument.

Tsimane'. Participants ($n = 64$, 31 female; mean age = 31.5 years, s.d. = 10.2 years) lived in a small rainforest village (Santa Maria) without electricity or tap water. Santa Maria could only be reached by canoeing up the Maniqui River (an indirect tributary of the Amazon). Age data are approximate, as many Tsimane' do not know their precise age in years[34]. Results of Fig. 2 were similar (all statistical comparisons produced the same result) if we limited the analysis to participants between the ages of 16 and 35, who were more likely to know their exact age ($n = 41$, mean age of 25.0 years, s.d. = 5.8 years). Music performance among the Tsimane' was largely limited to a small number of expert practitioners who would perform traditional songs at social gatherings; most Tsimane' participants did not perform music themselves.

Sample sizes for the United States were chosen based on previous knowledge of effect sizes and variability[21]. Sample sizes for Bolivia were chosen to approximately match the US groups (capital and town) or to be as large as possible given practical constraints (Tsimane'). See Extended Data Table 1 for a summary of the participant group demographics for Study 1.

Chord stimuli and analysis (Study 1). Chord stimuli were taken from a previous study[21]. Pleasantness ratings were obtained for 10 chord types: the minor second, major second, tritone, major seventh, and augmented triad (conventionally dissonant in Western music), and the major third, perfect fourth, perfect fifth, and major triad (conventionally consonant in Western music), along with the unison (conventionally neutral in Western music). The consonant and dissonant chords had been found previously to yield high and low pleasantness ratings, respectively, in Western listeners[21]. All chords were composed of intervals from the equal-tempered scale.

Synthetic chords were composed of synthetic notes: complex tones containing the first 10 harmonics (in sine phase). Harmonic amplitudes were attenuated at 14 dB per octave to mimic naturally occurring musical note spectra. Synthetic chords were given temporal envelopes that were the product of a half-Hanning window (10 ms) and a decaying exponential (decay constant of $2.5\,s^{-1}$) that was truncated at 2 s.

Sung chords were composed of notes (using the vowel /u/; 'ooh') produced and recorded by a professional singer (A. Lehr). They were high-pass filtered with a cutoff frequency of 150 Hz (fourth order Butterworth filter) to eliminate handling noise from the microphone. The temporal envelope of the onset of the sung notes was not altered apart from applying a half-Hanning window (10 ms). A linear ramp was applied over the last half of each note to fade them smoothly down to zero, such that the duration was the same as the synthetic notes. Example stimuli (taken from ref. 21) are available online: http://mcdermottlab.mit.edu/consonance_examples/index.html.

Each chord was presented four times, each time with a different root note, drawn from C#4, D#4, F4, and G4 for the synthetic chords and G#3, A#4, B#4, and D4 for the sung chords, where C4 denotes middle C (the range was lower for the sung chords to accommodate the range of the singer). There were thus a total of 40 chords presented for each of the synthetic or sung chord experiments. Sung and synthetic chords were presented in separate blocks. Tsimane' participants completed either the sung ($n = 33$) or the synthetic ($n = 31$) chord block due to time constraints. All other participants completed both blocks.

Ratings were averaged across consonant chords and dissonant chords to yield a single average rating for each aesthetic class for each participant. Apart from the analysis in Extended Data Fig. 1, all statistics were performed on these averaged ratings. The ratings of the unison were omitted from the main analysis to simplify data presentation (to two stimulus conditions per experiment).

Harmonic/inharmonic synthetic tone stimuli (Study 1). The harmonic and inharmonic tone stimuli were taken from a previous study[21]. The harmonic tones contained a subset of the frequencies of a normal harmonic tone. There were three types of harmonic tone: one containing harmonics 1, 2, 4 and 8, one containing harmonics 1, 2, 3, 5 and 9, and one containing just the first harmonic. There were also three types of inharmonic tone. Two of them were modifications of the second harmonic stimulus. In the first type of inharmonic tone (jittered), the harmonic frequencies were perturbed up (even components) or down (odd components) by 0.5 semitones. In the second type of inharmonic tone (shifted), all the harmonic frequencies were increased by 30 Hz. To avoid differences in beating between the harmonic and inharmonic tones, alternate frequency components were presented to opposite ears (the wide spacing between components additionally minimized beating). The third type of inharmonic tone consisted of two pure tones separated by 1.5 semitones, presented dichotically to minimize beating. The resulting stimulus was inharmonic, while being roughly matched in frequency to the single-frequency harmonic stimulus, but without introducing salient differences in beating.

Inharmonicity detection thresholds for a single mistuned frequency component in a harmonic complex have been previously estimated to be about 1% of the harmonic's frequency[26]. Our inharmonic perturbations were well above this: ~3% for the jittered condition, between ~1.5–10%, depending on the harmonic, for the shifted condition, and 9% for the dichotic pure tone stimulus. Each of the perturbations therefore produced audible inharmonicity, which to a typical Western listener is perceived as unpleasant[21].

All harmonic and inharmonic stimuli had components whose amplitudes decreased by 14 dB per octave to resemble naturally occurring sounds. All tones were given temporal envelopes that were the product of a half-Hanning window (10 ms) and a decaying exponential (decay constant of $2.5\,s^{-1}$) truncated at 2 s.

Each type of tone was presented four times, each time with a different base frequency, drawn without replacement from the 4-semitone range above middle C.

Smooth/rough synthetic tone stimuli and analysis (Study 1). The smooth and rough tone stimuli were taken from a previous study[21]. Tones that were perceptually rough or smooth were generated by presenting pairs of single frequencies to either the same or different ears (diotic and dichotic presentation, respectively). Diotic presentation of two similar but non-identical frequencies is known to produce the 'rough' sensation of beats, typically considered unpleasant by Western listeners[5,7,21,25]. In contrast, dichotic presentation of two such frequencies greatly attenuates perceived beats[35], but leaves the spectrum (and its pitch) unchanged relative to the diotic version[36]. The frequencies composing each stimulus were separated by either 0.75 or 1.5 semitones (1.5 for the low- and mid-frequency ranges, and 0.75 for the high-frequency range, to produce beat frequencies with prominent roughness), such that considerable beating was heard when presented diotically[5,7,21]. These stimuli, like the harmonic/inharmonic synthetic tones, were given envelopes that were the product of a half-Hanning window (10 ms) and a decaying exponential (decay constant of $2.5\,s^{-1}$) truncated at 2 s.

We generated smooth/rough tones in three frequency ranges. Each type of tone was presented four times, each time with a different base frequency, drawn without replacement from the 4-semitone range above either middle C (low-frequency), one octave above middle C (medium-frequency), or two octaves above middle C (high-frequency). The low-frequency dichotic tone also served as an inharmonic stimulus that could be compared to the single frequency harmonic stimulus.

The harmonic/inharmonic and smooth/rough tones were presented in random order in a single block. There were four exemplars of each of the eleven stimulus types, for a total of 44 trials. To analyse the harmonic/inharmonic tone data, the ratings of the three harmonic stimuli were averaged together, as were the ratings of the three inharmonic stimuli (one of which was the low-frequency dichotic

frequency pair stimulus). All statistics were performed on these two average ratings. To analyse the smooth/rough tone data, the ratings of the three dichotic (smooth) stimulus types were averaged together, as were the three diotic (rough) stimulus types. All statistics were performed on these two average ratings.

Vocalization stimuli (Study 1). Vocalizations were a subset of the Montreal Affective Vocalization Set[37], which consists of vocalizations produced by actors in a laboratory setting. Five vocalizations were selected from each of the categories of laughter, gasps, and crying, which had previously been found to be rated as pleasant (laughter) and unpleasant (gasps and cries) by North American listeners[22]. The stimuli were presented in random order in a single block of 15 trials. The crying sounds gave similar results to the gasps (low ratings by all groups) and their ratings were omitted from the analysis to simplify the data presentation (to two stimulus conditions per experiment).

Stimulus presentation (Study 1). All stimuli were normalized to have the same rms level. Stimuli were played out by a laptop computer over closed headphones (Sennheiser HD 280Pro). The volume was set to a level that was determined in pilot experiments to be comfortable (approximately 70 dB SPL), and that was fixed across participants.

Experimental protocol (Study 1). Sound stimuli for a given experiment were presented sequentially in random order. After each stimulus was presented, participants gave their rating verbally (participants could respond "like it a lot", "like it a little", "dislike it a little", or "dislike it a lot"), and the experimenter recorded the response on a data entry sheet. To avoid biasing participants' responses or data entry, experimenters were always blind to the stimulus being presented, and stimuli were audible only to the participant.

The experiments were completed in a single sitting along with several other experiments that are not reported here. Experiment order was counterbalanced across the participants in each group. Experiments were conducted in June–July 2011 (Tsimane'), January–February 2012 (San Borja and US musicians), July 2012 (La Paz), and March–April 2014 (US non-musicians).

Both studies were approved by the Tsimane' Council (the governing body of the Tsimane' in the Maniqui basin, where the studies took place), as well as the Committee on the Use of Humans as Experimental Subjects at MIT and the Committee for Protection of Human Subjects at Brandeis. Experiments were conducted with the informed consent of the participants.

Participants (Study 2). *US musicians.* A group of musically experienced US residents was tested to provide baseline preference measures. They were selected to have experience playing a musical instrument ($n=47$, 32 female; mean age $=25.9$ years, s.d. $=7.5$ years; mean number of years spent playing an instrument $=8.4$ years, s.d. $=6.5$ years, range $=1$–25 years). None of the participants were professional musicians; we label them here as 'musicians' primarily to distinguish them from the non-musicians of Study 1 who were selected to have almost no experience playing music or singing. All were residents of the Boston metropolitan area and had been born and raised in the United States. All participants completed all experiments in the study.

Tsimane'. Participants ($n=50$, 27 female; mean age $=27.4$ years, s.d. $=9.8$ years; as in Study 1, ages are approximate) lived in five small villages surrounding San Borja (Arenales, Manguito, Pachual, Limoncito, and Tacuaral). Unlike the village of Santa Maria from Study 1, the villages of Study 2 could be reached by automobile during the dry season and had begun to be equipped with electricity at the time of testing (some families were connected to the electrical grid). The choice of villages was dictated by accessibility given the local weather conditions during the study (roads were often muddy and impassable). We tested a larger group of participants and only analysed data from the subset who did not own a radio, to minimize the chance that they had significant exposure to Western music. Only 27 of these 50 participants completed the synthetic interval experiment (it was substituted for another experiment half-way through our fieldwork). One of the 50 participants did not complete the discrimination experiment. The size of the Tsimane' group was as large as possible given the practical constraints of the study (the US group was chosen to approximately match it in size). See Extended Data Table 2 for a summary of the participant group demographics for Study 2.

Chord stimuli (Study 2). Chord stimuli were generated as in Study 1. The synthetic and sung two-note chord experiments presented intervals from 0 to 11 semitones at each of 4 different root-pitches, yielding 48 trials in total. The sung triad experiment presented the major, minor, diminished, and augmented triads each at 4 different root-note pitches, yielding 16 trials in total (in the main analysis only the major and augmented triads were analysed, as they produced the largest contrast in US listeners). The root notes for the synthetic chords were the first four notes above middle C. For both sung chord experiments, the root notes were the first four notes above G#3 to accommodate the vocal range of the singer. The notes in the sung chord experiments were sung vowels ("aah" for two of the exemplars and "ooh" for the other two, randomly assigned to the root notes). Within experiments, trials were randomly ordered.

Vocal harmony stimuli (Study 2). Because the Tsimane' generally do not sing in groups and were unable to sing or play in coordination upon request, it was necessary to generate harmonies from solo recordings. We recorded two Tsimane' vocalists singing the same melodic phrase several times in a row. The phrases were chosen by the vocalists to be Tsimane' songs they knew well and tended to sing in social settings. We selected sets of phrases by each singer with very similar timing (by visually comparing the waveforms), yielding 26 phrases with two repetitions each. We then pitch-shifted one of the phrases in a set and added it to the other (unshifted) version of the phrase. Pitch shifting was performed using STRAIGHT[38]. The experiment presented harmonies with intervals ranging from 1 to 11 semitones, as well as single phrases and the unison (two different repetitions of the same phrase at the same pitch). There were 4 exemplars of each interval, each generated from a different melodic phrase, yielding 52 trials in total. Each of the 26 sung phrases was used twice in the experiment. We generated 15 random assignments of phrases to conditions with the constraint that the two occurrences of a phrase not be assigned to the same condition. One of these assignments was chosen at random for each participant. Although the song excerpts were not selected to be widely recognizable (they were simply those that the singers were comfortable singing), 19 out of the 50 participants were able to name at least one song that they recognized during the experiment.

Pure tone interval stimuli (Study 2). Pure tone intervals were generated by the same procedure used for the smooth/rough synthetic tone stimuli from Study 1, the only difference being that the tones in a pair were separated by integer numbers of semitones. The experiment presented intervals from 0 to 8 semitones (in diotic and dichotic versions), each at 2 different root-note frequencies (523 and 740 Hz). The diotic and dichotic 0-semitone conditions were identical.

Discrimination experiment stimuli/task (Study 2). Listeners heard harmonics 1–6 of a fundamental frequency (F0). On one-quarter of trials the even-numbered harmonics were mistuned by +12%, rendering the tone inharmonic and causing the even and odd harmonics to segregate and be heard (by typical Western listeners) as two separate tones. On another quarter of trials the even harmonics onset 150 ms earlier than the odd harmonics, again causing the even and odd harmonics to perceptually segregate. On the remaining 50% of trials the even harmonics were unperturbed and the complex tone tended to be heard as a single sound. The task was to judge whether a trial contained one or two sounds. After each trial the experimenter told the participant if they had answered correctly (the correct answer was considered to be "one" when the harmonics were unperturbed, and "two" when mistuned or offset). The tone F0 took 9 different values (0, 4 and 8 semitones above 200, 300 and 400 Hz), yielding 36 trials in total. Harmonics were 750 ms in duration and were given flat amplitude envelopes that began and ended with half-Hanning windows (10 ms). Stimuli were presented at 75 dB SPL.

The experiment was preceded by a practice experiment that was identical in all respects except that only 5 F0 s were used (200, 282.8, 300, 400 and 424.3 Hz), yielding 20 trials in total.

Stimulus presentation (Study 2). Same as in Study 1.

Experimental protocol (Study 2). Pleasantness ratings were obtained as in Study 1, except that the experimenter entered responses directly into the computer.

The experiments were completed in a single sitting. Experiment order was generated randomly for each participant, with the exception that the discrimination experiments were conducted first or last with equal probability (because the instructions were different than for the pleasantness rating experiments). Experiments were conducted in July 2015 (Tsimane') and July–October 2015 (US musicians).

Analysis. For preference experiments, the ratings for each participant were averaged across the exemplars for a condition (for example, laughs or gasps for the vocalization experiment, or major triad for one of the chord experiments), yielding one mean rating per condition per participant per experiment. In the chord, tone, and vocal harmony experiments, most analyses were done after averaging these mean ratings across groups of conditions, for example, the consonant chords/harmonies, or the harmonic tones, again yielding one mean rating for each participant for each group of conditions in each experiment. The results graphs plot these mean ratings averaged across participants, and the error bars plot the standard error of these mean ratings across participants.

For the discrimination experiment, we obtained one measure of sensitivity (d') for each condition for each participant from their hit and false alarm rates. The results graphs plot the mean and standard error of this sensitivity measure across subjects.

Statistics. Repeated-measures ANOVAs were used to test for main effects of stimulus condition and for interactions between the effect of stimulus condition and participant group. Mauchly's test was used to test for violations of the sphericity assumption. When Mauchly's test was significant, the Greenhouse–Geisser correction was used. Pair-wise comparisons were made with two-tailed t-tests, and were not corrected for multiple comparisons because they were preceded by

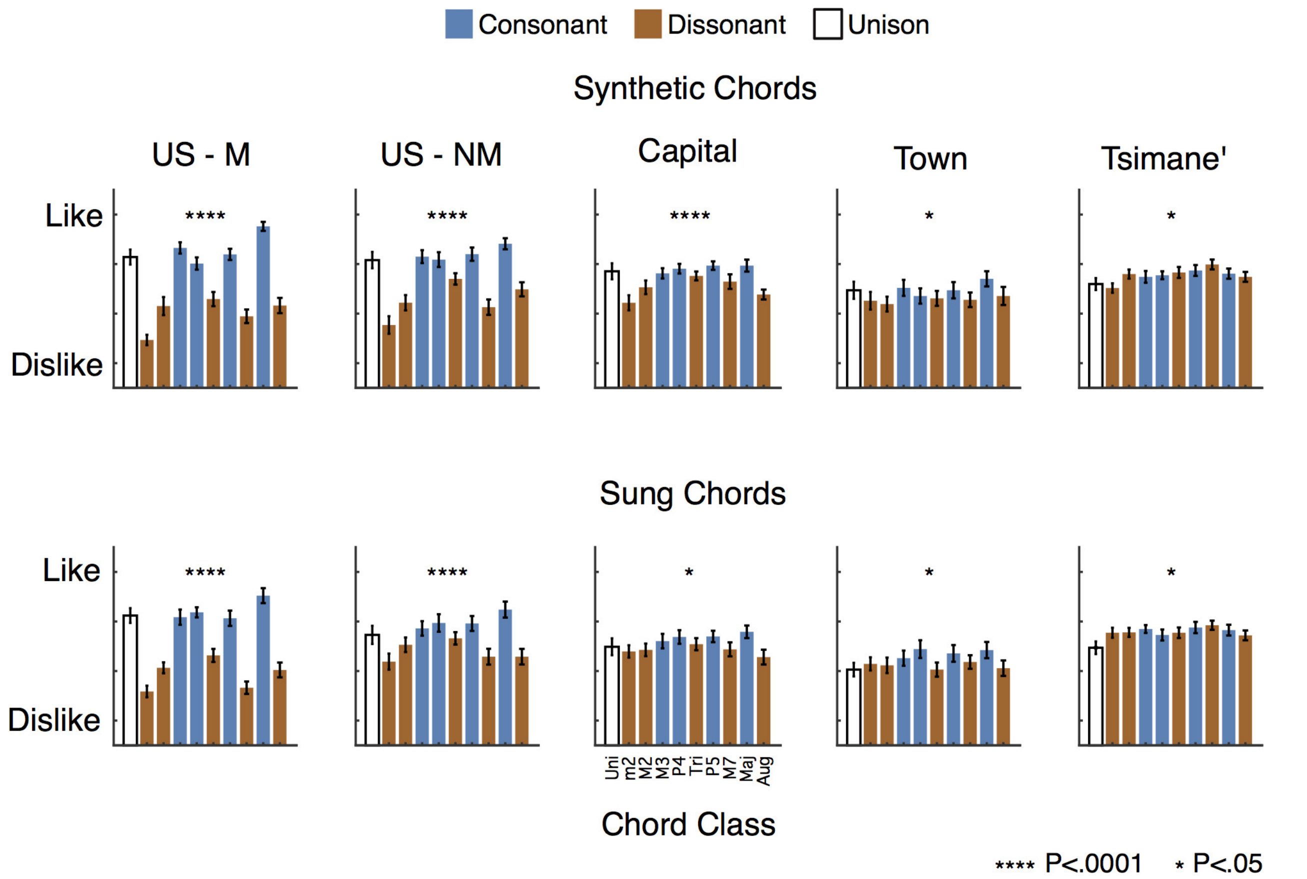
ANOVAs that revealed significant main effects or interactions. Pearson correlations were used to determine the relationship between chord ratings and interval size (Extended Data Fig. 2) and to test the reliability of ratings of song excerpts and vocal harmonies (Extended Data Fig. 3). Data distributions were assumed to be normal, and were evaluated as such by eye. The exceptions to the above were in the analysis of individual melodic phrase ratings in the vocal harmony experiment (Extended Data Fig. 3) and of the pure tone interval experiment from Study 2 (Fig. 3f), in which only two ratings per condition were collected per subject due to the large number of conditions (producing non-normal data distributions), necessitating non-parametric tests. In these cases, Friedman's non-parametric test of differences among repeated measures was used to test for main effects, and the Wilcoxon signed rank test was used for pairwise comparisons.

Tsimane' music. Tsimane' music consists of songs that have apparently been passed down for generations. At present these songs are sung primarily at social gatherings by particular individuals who know the songs, typically when adults gather to consume "chicha", a traditional homemade alcoholic beverage. The primary documentation of Tsimane' music comes from the work of J. Riester and G. Rocckel (an anthropologist and ethnomusicologist, respectively), who compiled 140 traditional songs and analysed their subject matter and musical content[20]. We have also compiled Tsimane' songs as part of a broader longitudinal study[39]. Known Tsimane' songs tend to reflect their everyday lives, concerning production activities (such as hunting), division of labour, love, social organization, and religion. Musically, many songs are based on anhemitonic pentatonic scales. Others are constructed from three notes spanning a perfect fourth (typically divided into intervals of a whole tone and minor third). Examples of recorded songs are provided in Supplementary Information (Supplementary Audio 1 and 2). Instrumental renditions of songs are also common, typically on wind or string instruments. Drums exist but do not seem to be used to accompany other instruments or singing, at least not any more. There is one report of Tsimane' performing traditional dances in the mid-twentieth century[40], but to the best of our knowledge, traditional dancing is no longer practiced.

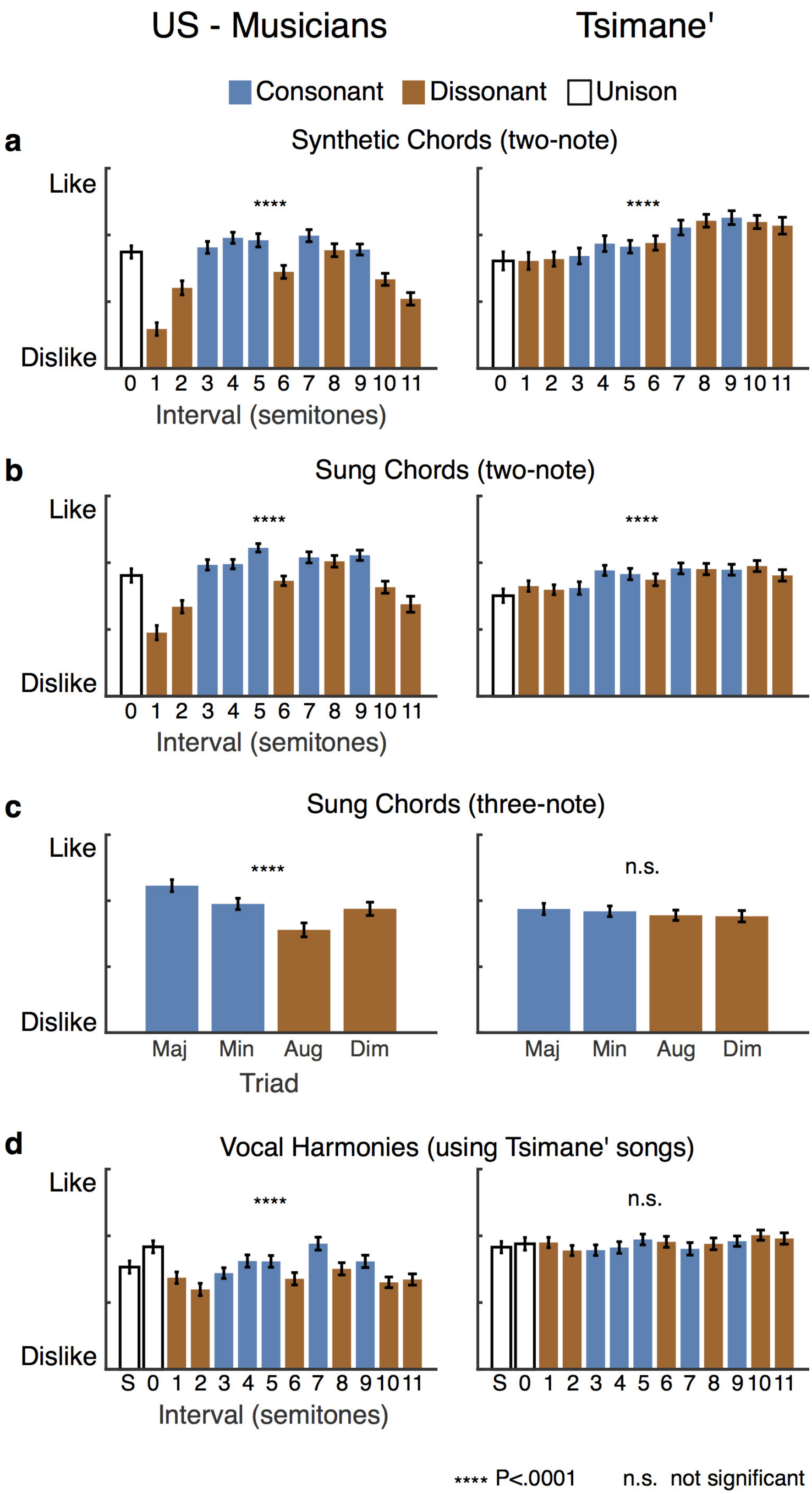
For our purposes, the most interesting feature of Tsimane' music is that group performances appear to be absent. We substantiated this over several days of interviews with 10 Tsimane' musicians. They consistently reported that musical performances often occur in social settings, but that only one person performs at a time. Multiple musicians might perform in succession, but not concurrently; instrumental accompaniment to singing is rare. We were initially sceptical of this claim, and so brought pairs of musicians together and asked them to perform together. They were usually reluctant to do so (despite being eager to perform solo songs for each other), and on the few occasions when we could elicit concurrent performances, they were unable to coordinate. Our experience suggests that group musical performance in Tsimane' culture is rare at best.

The origins of the Tsimane' music tradition are unknown, but they clearly have had some contact with the music traditions of other cultures for at least the past century, including exposure to Western musical instruments, primarily through friars and missionaries as well as through Bolivian highlanders. It is also clear that the Tsimane' musical tradition is slowly fading from their culture, in part due to the influence of missionaries. For instance, shamans were once a staple guardian of Tsimane' songs, but were banned by Protestant missionaries who became influential in parts of the Tsimane' territory[41]. The past two decades have also seen the diffusion of highland Andean music and Christian hymns into the Tsimane' territory through the radio station of Protestant missionaries, nearby commercial radio stations, and highland migrants moving into the lowlands. As part of a longitudinal study with the Tsimane we found that of the 2,549 adults interviewed annually during 2002–2010 in 13 villages along the Maniqui River (varying in proximity to market towns and roads), 30% reported owning at least one battery-operated radio. These figures may overstate the impact of radios because some of the radios lacked batteries, or were broken. It is also unclear to what extent the Tsimane' use their radios to listen to music, as they are also a primary source of local news. More recently, the Bolivian government has started to expand the electricity grid to the countryside surrounding the town of San Borja, and as a result, some villages have begun to acquire electricity. Our impression is that the effect of these outside influences has been to reduce the prevalence of traditional Tsimane' music in their culture. Most Tsimane' that we interviewed had knowledge of the traditional songs, but this might well diminish in the future.

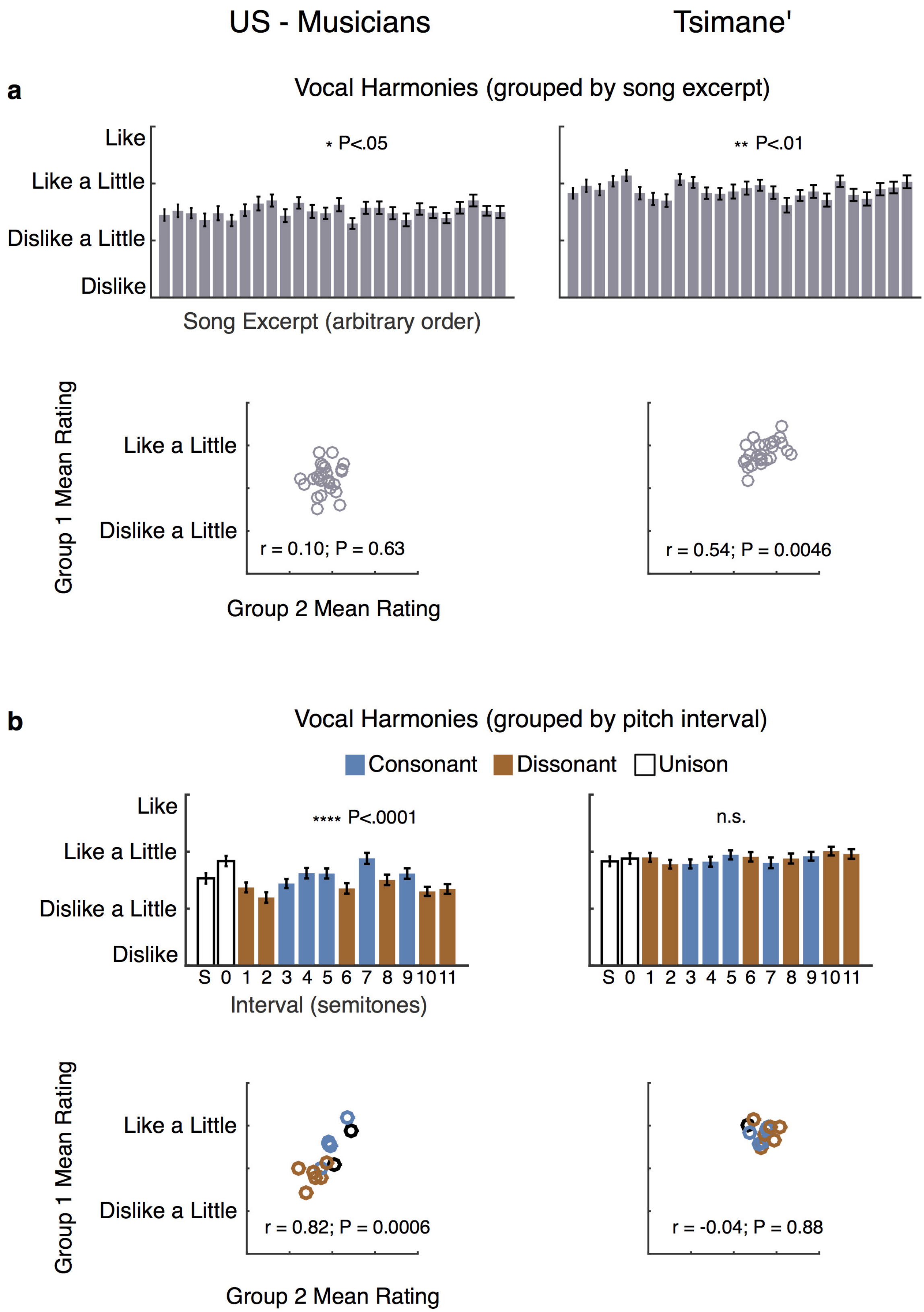
31. Malmberg, C. F. The perception of consonance and dissonance. *Psychol. Monogr.* **25,** 93–133 (1918).
32. Hutchinson, W. & Knopoff, L. The acoustical component of western consonance. *Interface* **7,** 1–29 (1978).
33. Roberts, L. Consonance judgments of musical chords by musicians and untrained listeners. *Acustica* **62,** 163–171 (1986).
34. Gurven, M., Kaplan, H. & Supa, A. Z. Mortality experience of Tsimane Amerindians of Bolivia: regional variation and temporal trends. *Am. J. Hum. Biol.* **19,** 376–398 (2007).
35. Rutschmann, J. & Rubinstein, L. Binaural beats and binaural amplitude-modulated tones: successive comparison of loudness fluctuations. *J. Acoust. Soc. Am.* **38,** 759–768 (1965).
36. Bernstein, J. G. & Oxenham, A. J. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* **113,** 3323–3334 (2003).
37. Belin, P., Fillion-Bilodeau, S. & Gosselin, F. The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* **40,** 531–539 (2008).
38. Kawahara, H. & Morise, M. TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *Sadhana* **36,** 713–722 (2011).
39. Leonard, W. R. *et al.* The Tsimane' Amazonian Panel Study (TAPS): Nine years (2002-2010) of annual data available to the public. *Econ. Hum. Biol.* **19,** 51–61 (2015).
40. Hissink, K. Felsbilder und salz der Chimanen-Indianer. *Paideuma (Wiesb.)* **6,** 60–68 (1955).
41. Huanca, T. *Tsimane' Oral Tradition, Landscape, and Identity in Tropical Forest* (Wa-Gui, 2008).



Extended Data Figure 1 | Average ratings of individual chords from the synthetic and sung chord experiments from Study 1. Ten chords were presented: the unison, minor second, major second, major third, perfect fourth, tritone, perfect fifth, major seventh, major triad, and augmented triad. The composite ratings plotted in Fig. 2 were averages of those for the consonant (blue) and dissonant (brown) chords. Ratings are from 23 US musicians (US-M), 25 US non-musicians (US-NM), 24 Bolivian city-dwellers (capital), 26 Bolivian town-dwellers, and 64 Tsimane'. Asterisks denote statistical significance of a repeated-measures ANOVA across all chord ratings. Data are mean and s.e.m.

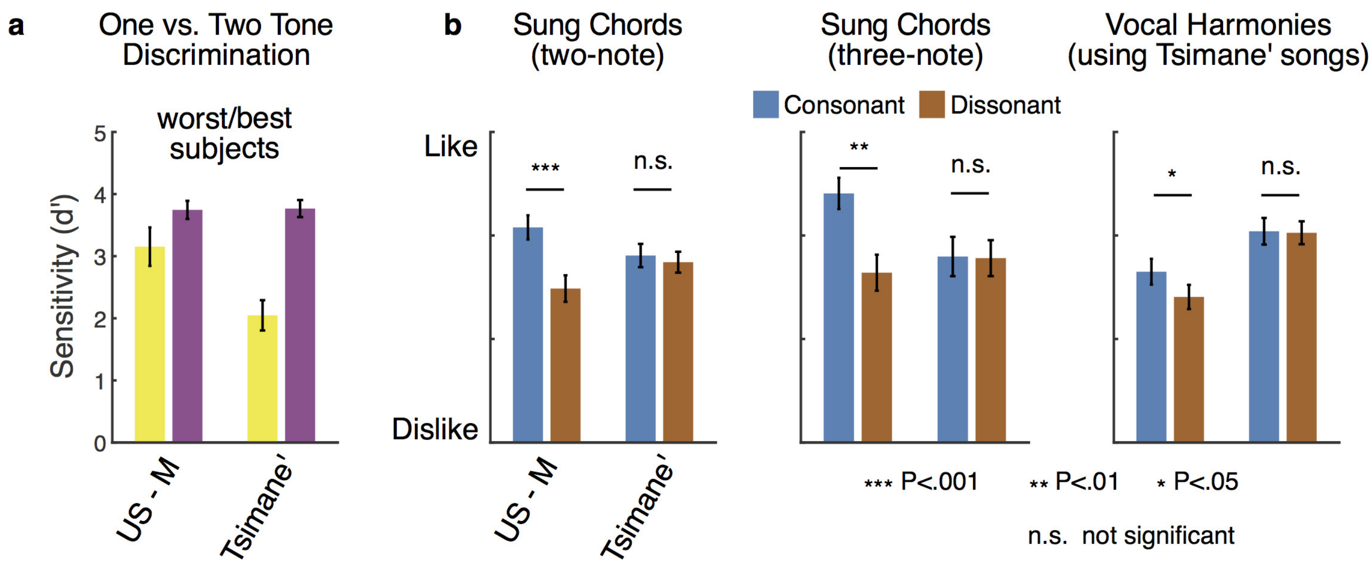


Extended Data Figure 2 | Average ratings of individual chords and vocal harmonies from Study 2. Rating variation for US listeners ($n = 47$) was largely determined by consonance and dissonance, whereas for Tsimane' listeners ($n = 50$) it was largely driven by interval size (in the two-note chord experiments in which they exhibited significant variation). 'S' denotes a single vocal phrase in the vocal harmonies experiment, whereas '0' denotes the unison condition (two concurrently presented exemplars of the same phrase at the same pitch). Data are mean and s.e.m.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Pleasantness ratings of vocal harmonies from Study 2, averaged across pitch intervals or across song excerpts, for US musicians and Tsimane'. a, Ratings of song excerpts used to create harmonies (averaged across intervals). Each participant heard each excerpt twice, in each of two different randomly assigned interval conditions. The ratings of the two excerpt occurrences were averaged for each participant (47 US musicians and 50 Tsimane'). Top panels plot these mean ratings averaged across all participants. Asterisks denote statistical significance of Friedman's non-parametric test of differences among repeated measures across song excerpts. Ratings of US listeners varied across song excerpts ($\chi^2(25) = 40.92$, $P = 0.02$), as did those of Tsimane' listeners ($\chi^2(25) = 49.01$, $P = 0.003$), but the preferences of the two participant groups were not significantly correlated ($r = -0.19$, $P = 0.35$). Bottom panels plot the mean ratings of each excerpt averaged across the first and last halves of the participants tested. The mean song excerpt ratings of the two half-groups were significantly correlated for Tsimane' but not for US participants, indicating that the preferences in US listeners were not reliable. **b**, Ratings of pitch intervals of the harmonies (averaged across song excerpts). Each participant heard each pitch interval four times, each time generated with a different song excerpt. The ratings of the four interval occurrences were averaged for each participant. Top panels (replicated from Extended Data Fig. 2) plot these mean ratings averaged across all participants. Asterisks denote statistical significance of a repeated-measures ANOVA across all chord ratings. Bottom panels plot the mean rating of each pitch interval averaged across the first and last halves of the participants tested. The mean interval ratings of the two groups were significantly correlated for US but not for Tsimane' participants. Data are mean and s.e.m.



Extended Data Figure 4 | Discrimination and preference data for worst US and best Tsimane' participant subsets from Study 2. **a**, Discrimination performance for the worst US-M participants (bottom third, $n = 16$) and best Tsimane' participants (top third, $n = 16$), selected based on performance in the onset asynchrony condition. This selection criterion produced a group of Tsimane' listeners who achieved an average d' of 2.0 on the mistuned condition—good performance in absolute terms, and only slightly worse than the poorly performing Western listeners. **b**, Pleasantness ratings for the subsets of participants from **a** for conventionally consonant and dissonant sung chords and vocal harmonies. This subset of Tsimane' listeners remained indifferent to dissonance, rating consonant and dissonant sung intervals, triads, and harmonies similarly ($t(15) < 1.2$, $P > 0.28$ in all cases). By contrast, the Western subset yielded significant consonant preferences ($t(15) > 2.65$, $P < 0.05$ in all cases). These results suggest that the absence of a consonance preference cannot be explained by a lack of sensitivity to the underlying stimulus distinction—the Tsimane' were able to distinguish harmonic from inharmonic tones despite not having a preference for one over the other. Data are mean and s.e.m.

Extended Data Table 1 | Summary of participant group demographics for Study 1

Participant Summary Table for Study 1

	US (Musicians)	US (Non-musicians)	La Paz (Capital)	San Borja (Town)	Santa Maria (Tsimane' village)
N total	23	25	24	26	64
N female	19	12	14	15	31
Mean age	20.2	37.9	29.9	31.6	31.5

Extended Data Table 2 | Summary of participant group demographics for Study 2

Participant Summary Table for Study 2

	US (Musicians)	Tsimane'
N total	47	50
N female	32	27
Mean age	25.9	27.4

LETTER

doi:10.1038/nature18928

Transfer of mitochondria from astrocytes to neurons after stroke

Kazuhide Hayakawa[1], Elga Esposito[1], Xiaohua Wang[1,2], Yasukazu Terasaki[1], Yi Liu[1,2], Changhong Xing[1], Xunming Ji[2] & Eng H. Lo[1]

Neurons can release damaged mitochondria and transfer them to astrocytes for disposal and recycling[1]. This ability to exchange mitochondria may represent a potential mode of cell-to-cell signalling in the central nervous system. Here we show that astrocytes in mice can also release functional mitochondria that enter neurons. Astrocytic release of extracellular mitochondrial particles was mediated by a calcium-dependent mechanism involving CD38 and cyclic ADP ribose signalling. Transient focal cerebral ischaemia in mice induced entry of astrocytic mitochondria into adjacent neurons, and this entry amplified cell survival signals. Suppression of CD38 signalling by short interfering RNA reduced extracellular mitochondria transfer and worsened neurological outcomes. These findings suggest a new mitochondrial mechanism of neuroglial crosstalk that may contribute to endogenous neuroprotective and neurorecovery mechanisms after stroke.

Astrocytes have broad roles in the central nervous system, and are involved in the regulation of neurodevelopment, neurotransmission, cerebral metabolism and blood flow[2–4]. Normal astrocytes protect neurons against oxidative stress and excitotoxicity[5–7]. By contrast, unhealthy astrocytes may release deleterious factors that damage neurons[8,9]. Healthy mitochondria may be essential for these neuroglial protective mechanisms because inhibition of astrocytic mitochondria makes neurons vulnerable to cell death[10]. Mitochondria comprise the intracellular cores for energetics and viability[11], but under some conditions mitochondria might also be released into the extracellular space[12]. For example, retinal neurons may transfer mitochondria to astrocytes for disposal and recycling[1], and bone-marrow-derived stromal cells may transfer mitochondria into pulmonary alveoli to suppress acute lung injury[13].

Here we investigated whether astrocytes can produce functional extracellular mitochondria to support neuronal viability after ischaemic stroke. Electron microscopy confirmed the presence of extracellular particles containing mitochondria in conditioned media from rat cortical astrocytes (Fig. 1a, Extended Data Fig. 1a). qNano analysis revealed that astrocyte-derived mitochondrial particles isolated by fluorescence-activated cell sorting (FACS) spanned a range of sizes from 300 to 1,100 nm (Extended Data Fig. 1b–d), and included populations that were positive for β1-integrin (79%) and CD63 (43%) (Extended Data Fig. 2). MitoTracker labelling suggested that these extracellular mitochondria may still be functional (Fig. 1b), and filtration of astrocyte-derived conditioned media through 0.2-μm filters depleted the amounts of functional mitochondria and reduced measurements of mitochondrial ATP, membrane potential and oxygen consumption (Fig. 1b–e).

An important question is whether extracellular mitochondria represent active signals or merely cellular debris. To address this, we investigated whether stimulated astrocytes could actively produce extracellular mitochondria. CD38 catalyses the synthesis of a calcium messenger, cyclic ADP-ribose (cADPR), in mitochondrial membranes[14,15]. In the brain, CD38 is mainly expressed in glial cells, and may have a role in neuroglial crosstalk since astrocytes increase CD38 expression in response to glutamate release from neurons[16]. On the basis of this background literature and the fact that most actively secreted cellular events involve calcium regulation, we decided to assess CD38–cADPR–calcium signalling as a candidate mechanism for the astrocytic production of extracellular mitochondria. First, we confirmed that rat cortical astrocytes expressed CD38 protein and contained CD38 or ADPR cyclase activity (Fig. 1f, g). We then used two methods to modify this pathway. When astrocytic CD38 was upregulated using CRISPR/Cas9 activation plasmids, the functional endpoints of extracellular mitochondria were significantly increased in conditioned media (Fig. 1h–k). When astrocytes were stimulated by cADPR to activate CD38 signalling, extracellular mitochondria were increased in conditioned media along with enhancement of functional endpoints in a calcium-dependent manner (Fig. 1l–n, Extended Data Fig. 3). Stimulation with cADPR did not appear to damage astrocyte viability (Fig. 1o), suggesting that this release of extracellular mitochondria was not due to nonspecific cytotoxicity.

If astrocytes can produce functional extracellular mitochondria, it is possible that these signals could affect adjacent neurons. When rat cortical neurons were subjected to oxygen–glucose deprivation, intracellular ATP levels fell and neuronal viability decreased, as expected (Fig. 2a–c, Extended Data Fig. 4a). When astrocyte-derived conditioned media containing extracellular mitochondrial particles was added to neurons, ATP levels increased and neuronal viability recovered (Fig. 2a–c, Extended Data Fig. 4a). However, when extracellular mitochondria were removed from the astrocyte-conditioned media, neuroprotection was no longer observed (Fig. 2a–c, Extended Data Fig. 4a). Similar results were obtained with immunostaining-based cell counts (Fig. 2d). As a control, ATP-liposomes were not significantly protective (Fig. 2e), suggesting that the entry of astrocytic mitochondria into neurons may generate additional benefits beyond ATP energetics per se. Fluorescent microscopy confirmed that astrocyte-derived mitochondria were present within treated neurons (Fig. 2f and Extended Data Fig. 4b).

Beyond the prevention of acute neuronal death, delayed neuroplasticity is also important for stroke outcomes. CD38 may be important for brain plasticity because CD38-deficient mice show worsened recovery after brain injury[17] and CD38 mutations may comprise risk factors for behavioural dysfunction[18]. Hence, we asked whether CD38-mediated astrocyte-to-neuron mitochondrial transfer may also influence neuroplasticity. Neurons were labelled with the green fluorescent protein (GFP)-containing construct CellLight Mitochondria-GFP and astrocytes were separately labelled with the red fluorescent probe MitoTracker Red CMXRos, and the two cell types were co-cultured together for 24 h. Confocal microscopy indicated that astrocyte-derived mitochondria were detected within soma and axon (Fig. 3a), and in these co-culture conditions, astrocytes supported neuronal survival after serum/glucose starvation in a CD38-dependent manner (Extended Data Fig. 5). When astrocytic mitochondria

[1]Neuroprotection Research Laboratory, Departments of Radiology and Neurology, Massachusetts General Hospital and Harvard Medical School, Charlestown, Massachusetts 02129, USA.
[2]Cerebrovascular Research Center, Xuanwu Hospital, Capital Medical University, Beijing 100053, China.

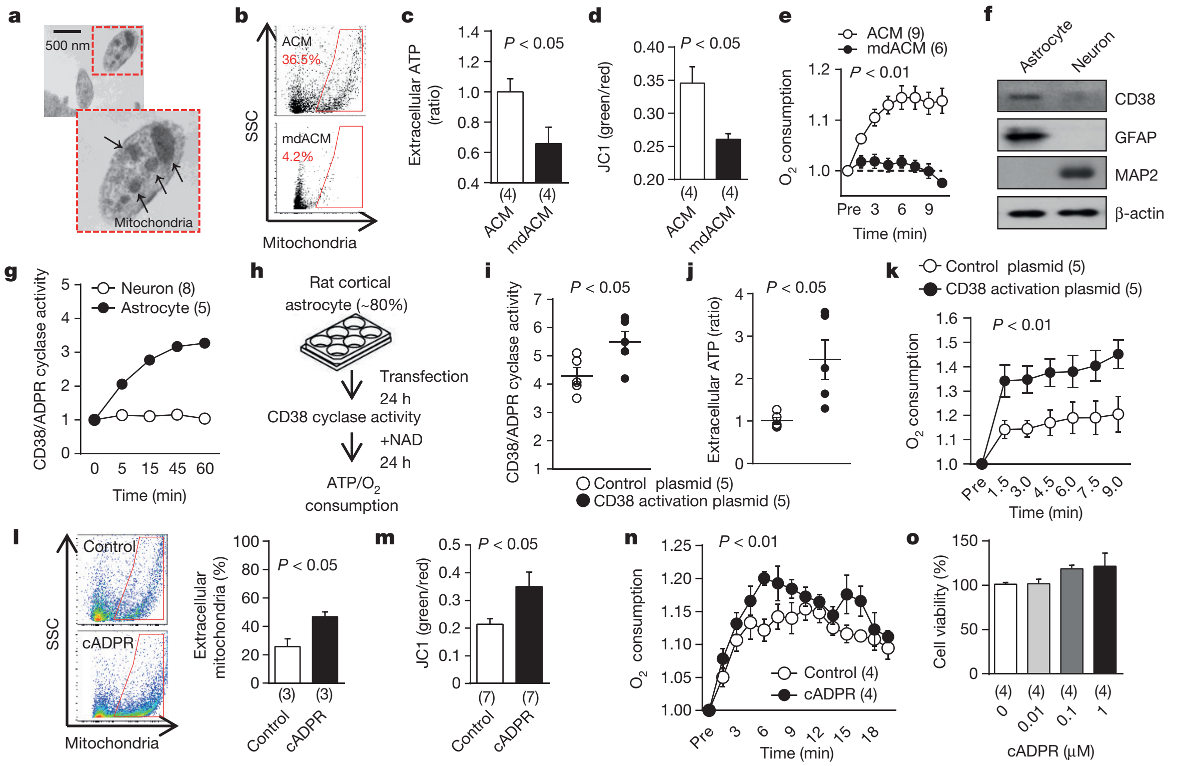


Figure 1 | Astrocytic CD38 and extracellular mitochondria. **a**, Transmission electron microscopy (TEM) of extracellular mitochondria in astrocyte-derived conditioned medium (ACM). Scale bar, 500 nm. **b**, Rat cortical astrocytes were labelled with MitoTracker Red CMXRos. FACS showed that filtration through 0.2-μm filters depleted levels of extracellular mitochondria in ACM (mitochondria-depleted ACM; mdACM). **c**–**e**, 0.2-μm filters reduced several markers of extracellular mitochondrial function in ACM: extracellular ATP (**c**, $n = 4$ independent experiments), membrane potential (JC1 ratio; **d**, $n = 4$ independent experiments), and oxygen consumption (**e**, $n = 3$ biological replicates, $n = 9$ or 6 independent experiments). **f**, Western blot confirmed higher CD38 levels in rat cortical astrocytes than in neurons. **g**, High and low levels of CD38 cyclase activity in astrocytes and neurons, respectively ($n = 2$ biological replicates, $n = 8$ or 5 independent experiments). **h**, Experimental schematic for testing CRISPR/Cas9-mediated CD38 activation. NAD, β-nicotinamide adenine dinucleotide sodium salt (200 μM). **i**, Twenty-four hours after transfection, CD38 cyclase activity was upregulated by CD38 activation plasmid ($n = 2$ biological replicates, $n = 5$ independent experiments). **j**, **k**, Extracellular ATP production (**j**) and oxygen consumption (**k**) were significantly increased by CD38 activation ($n = 2$ biological replicates, $n = 5$ independent experiments). **l**, FACS showed that extracellular mitochondria were increased by cADPR (1 μM) stimulation in astrocytes ($n = 3$ independent experiments). **m**, cADPR (1 μM) increased extracellular mitochondria membrane potential at 24 h ($n = 2$ biological replicates, $n = 7$ independent experiments). **n**, Oxygen consumption in extracellular mitochondria was increased by cADPR ($n = 2$ biological replicates, $n = 4$ independent experiments). **o**, cADPR did not cause astrocyte toxicity ($n = 2$ biological replicates, $n = 4$ independent experiments). Data are mean ± s.e.m. *P* values are from an unpaired *t*-test (**c**, **d**, **i**, **j**, **l**, **m**) or two-way analysis of variance (ANOVA) (**e**, **k**, **n**).

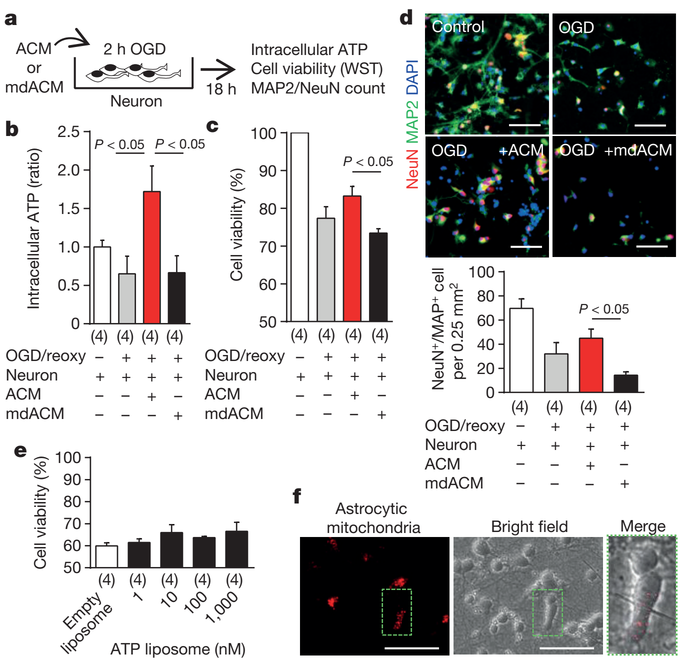


Figure 2 | Astrocytic extracellular mitochondria and neuroprotection. **a**, Experimental schematic to test neuroprotective effects of ACM or mdACM against oxygen–glucose deprivation (OGD) in rat cortical neurons. WST, water-soluble tetrazolium salt dye for cell viability assay. **b**, ACM but not mdACM rescued ATP levels in damaged neurons ($n = 2$ biological replicates, $n = 4$ independent experiments). **c**, ACM but not mdACM recovered neuronal viability after OGD ($n = 2$ biological replicates, $n = 4$ independent experiments). **d**, Immunostaining confirmed that neuroprotective effect of ACM but not mdACM ($n = 2$ biological replicates, $n = 4$ independent experiments). NeuN and MAP2 are neuronal markers. Scale bars, 100 μm. **e**, No statistically significant neuroprotection with liposomal ATP (1–1,000 nM) after OGD. **f**, Fluorescent microscopy suggests the presence of astrocyte mitochondria (labelled with MitoTracker Red CMXRos, 200 nM) within neurons. Scale bars, 100 μm. Data are mean ± s.e.m. *P* values are from a one-way ANOVA followed by Tukey's test.

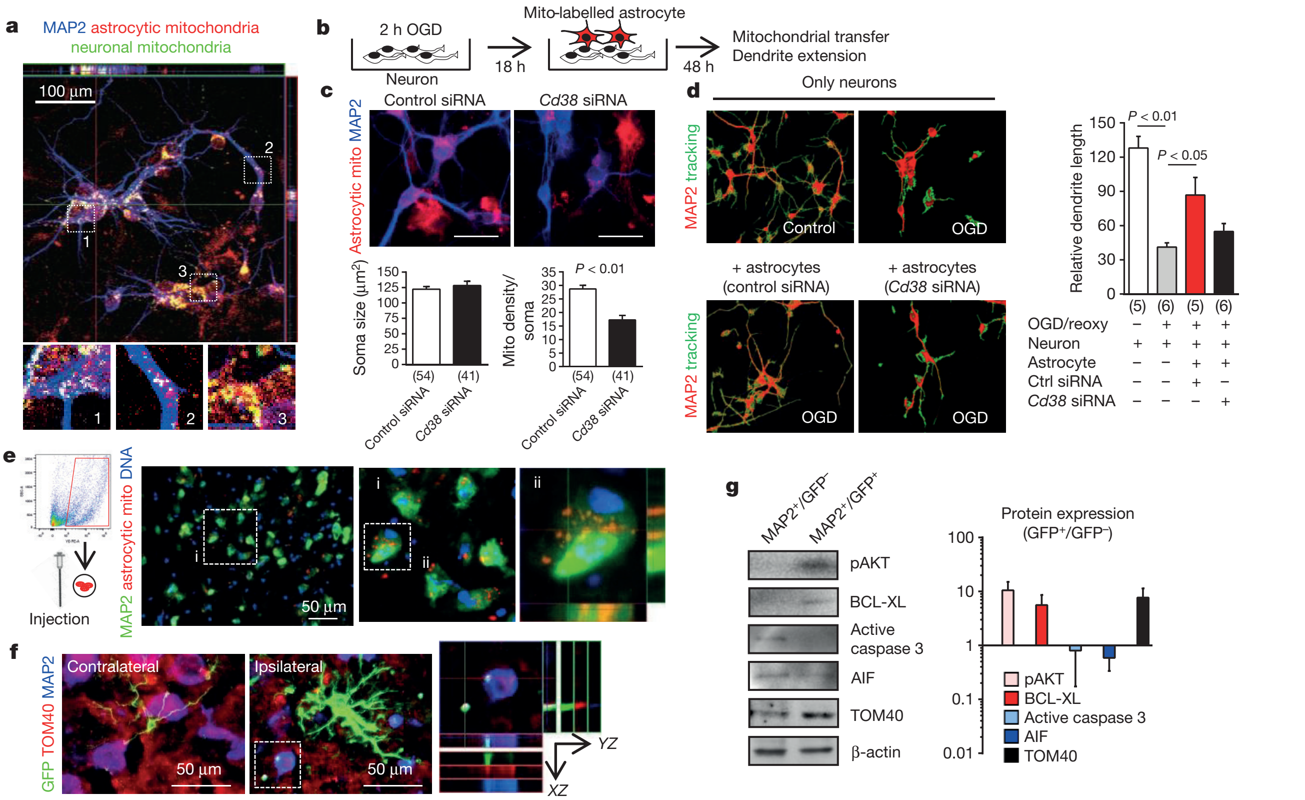


Figure 3 | Astrocytic mitochondria and neuroplasticity after ischaemic stress. a, Confocal microscopy revealed that astrocytic mitochondria (red, MitoTracker Red CMXRos) may be transferred into neural soma (1) and axon (2), and some may fuse with neuronal mitochondria (3; green, CellLight Mitochondria-GFP). **b**, Experimental schematic for co-culture studies. **c**, Soma size was unchanged but astrocytic mitochondrial density in neuronal soma was significantly decreased when CD38 was suppressed in astrocytes ($n = 54$ or 41 soma from $n = 2$ biological replicates, $n = 3$ independent experiments were counted). Scale bars, 20 μm. **d**, Quantification of dendrite elongation (MAP2 staining) ($n = 2$ biological replicates, $n = 5$ or 6 independent experiments). **e**, Male C57BL/6 J mice were subjected to 60 min transient focal ischaemia. Three days later, astrocyte mitochondrial particles (1,000 particles per 2 μl, MitoTracker Red CMXRos) were infused into cerebral cortex. Z-stack images showed transplanted astrocytic mitochondria (red) within peri-infarct neurons at 24 h. **f**, FVB/N-Tg (GFAPGFP)14Mes/J transgenic mice with fluorescently labelled astrocytes were subjected to 30 min transient focal ischaemia. Immunohistochemistry at 24 h suggested that GFP (GFAP)-positive particles co-stained with mitochondrial TOM40 were present in MAP2-positive neurons in peri-infarct cortex. **g**, Western blot indicated that GFP-positive neurons upregulated cell-survival-related proteins (phosphorylated-AKT, BCL-XL) but not apoptosis-related proteins (caspase 3, AIF) along with an increase in mitochondrial TOM40 ($n = 3$ mice). Isolated neurons expressed mature (neurofilament) but not neural stem-cell markers (nestin) (see Extended Data Fig. 7f). Data are mean ± s.e.m. P values are from an unpaired t-test (**c**) or one-way ANOVA followed by Tukey's test (**d**).

were made dysfunctional via inhibition of mitochondrial aconitase, cADPR-stimulated astrocytes no longer supported neuronal survival and axonal extension (Extended Data Fig. 6). To assess our hypothesis further, we asked whether this ability of astrocytes to transfer mitochondria could in fact enhance neuroplasticity under pathological conditions. Control or CD38-silenced astrocytes were co-cultured with surviving neurons after oxygen–glucose deprivation (Fig. 3b). Short interfering RNA (siRNA) suppression of CD38 in astrocytes reduced mitochondrial transfer (Fig. 3c) and dendrite regrowth after injury (Fig. 3d).

Taken together, these cellular findings appear consistent with the overall hypothesis that CD38 signalling may help astrocytes to transfer mitochondria into neurons and promote survival and plasticity after injury. But it is unclear whether this mechanism works *in vivo*. To address this question, we turned to a mouse model of focal cerebral ischaemia. First, primary mouse cortical astrocyte cultures were labelled with MitoTracker Red CMXRos and extracellular mitochondrial particles were collected. Mice were then subjected to focal cerebral ischaemia, and 3 days later, extracellular mitochondrial particles were directly injected into the peri-infarct cortex. After 24 h, immunostaining suggested that transplanted astrocytic mitochondria were present in neurons (Fig. 3e). Next, we turned to FVB/N-Tg (GFAPGFP)14Mes/J transgenic mice in which astrocytes are fluorescently labelled. When these mice were subjected to focal cerebral ischaemia, fluorescent mitochondrial particle signals appeared within adjacent neurons at 24 h after stroke (Fig. 3f). Neurons that were collected from ischaemic peri-infarct cortex via flow cytometry showed a general upregulation of cell-survival-related signals such as phosphorylated AKT and BCL-XL, along with an increase in the mitochondria marker TOM40 (Fig. 3g, Extended Data Fig. 7).

Finally, we attempted loss-of-function experiments to determine whether blocking CD38 signalling results in worsened outcomes after stroke. In our mouse model of focal cerebral ischaemia, CD38 was upregulated in the peri-infarct cortex (Extended Data Fig. 8a). At 5 days after stroke, *Cd38* siRNA or control siRNA was injected into cerebral ventricles (Fig. 4a). By 2 days after siRNA injections, total CD38 expression in the peri-infarct cortex was successfully downregulated (Extended Data Fig. 8b). There were no clear differences in infarct area or the total levels of GFAP-positive reactive astrocytes (Fig. 4b, c), but astrocyte subsets that expressed CD38 were significantly decreased without affecting the number of other CD38-expressing cells such as CD8 T cells and microglia/macrophages[19] (Fig. 4c, Extended Data Fig. 8c–g). To assess the levels of extracellular mitochondrial particles in this *in vivo* model, flow cytometry was used to analyse cerebrospinal fluid (CSF). GFAP-positive mitochondria were detected in CSF, and *Cd38* siRNA injections appeared to reduce

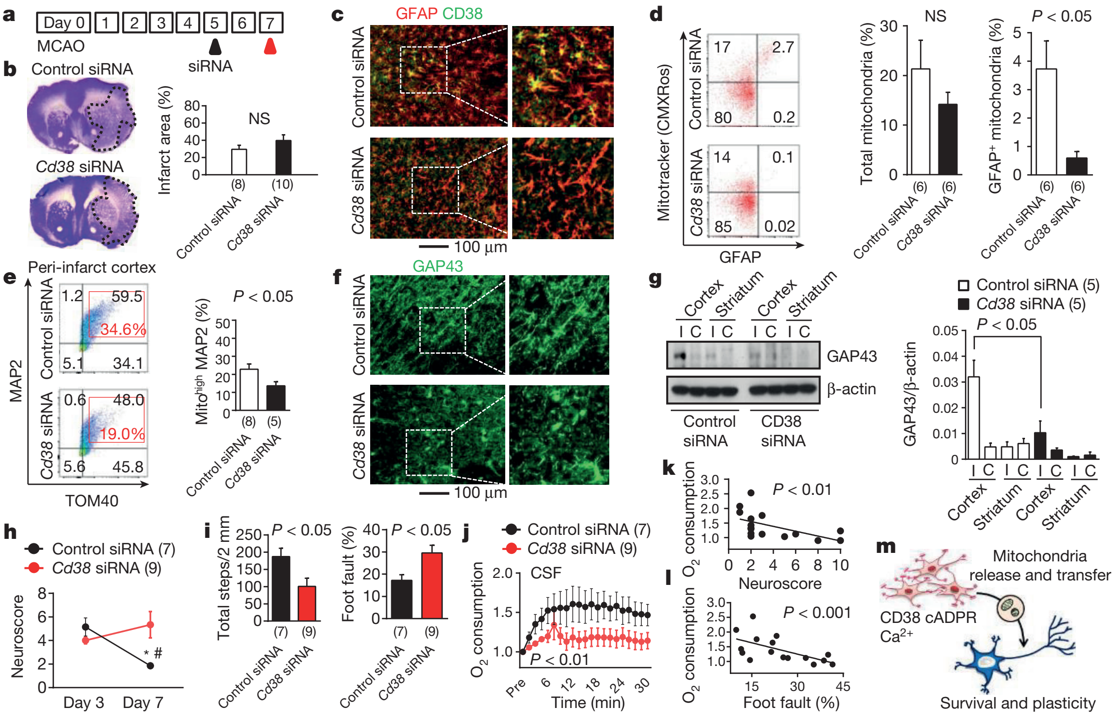


Figure 4 | Effects of *Cd38* siRNA on focal cerebral ischaemia. a, Male C57BL/6 J mice were subjected to transient 60 min focal ischaemia, and control or *Cd38* siRNA was injected into lateral ventricles 5 days after stroke. MCAO, middle cerebral artery occlusion. **b**, Nissl staining showed no difference in infarct size ($n = 8$ or 10 mice per group). NS, not significant. **c**, Immunostaining demonstrated that astrocytic CD38 was diminished by *Cd38* siRNA. **d**, Astrocytic CD38 suppression by siRNA reduced GFAP-positive mitochondria in CSF at 7 days ($n = 6$ mice per group). **e**, Neuronal mitochondria were decreased by *Cd38* siRNA ($n = 8$ or 5 mice per group). **f**, *Cd38* siRNA attenuated peri-infarct GAP43 immunostaining. **g**, Western blot confirmed a reduction of peri-infarct GAP43 protein within *Cd38*-siRNA-treated brains ($n = 5$ mice per group). C, contralateral hemisphere; I, ipsilateral hemisphere. **h**, **i**, Suppression of CD38 signalling worsened neurological outcomes in neurological severity scores (neuroscore) (**h**) and grid-walking test (**i**) ($n = 7$ or 9 mice per group). *$P < 0.05$ versus day 3 control siRNA; #$P < 0.05$ versus day 7 *Cd38* siRNA. **j**, CD38 suppression decreased oxygen consumption in CSF mitochondria ($n = 7$ or 9 mice per group). **k**, **l**, Mitochondrial function in CSF was negatively correlated with neurological outcomes. Data are mean ± s.e.m. *P* values are from an unpaired *t*-test (**b**, **d**, **e**, **i**), one-way ANOVA followed by Tukey's test (**g**), repeated-measures two-way ANOVA followed by Bonferroni test (**h**), two-way ANOVA (**j**), or Spearman's rank correlation (**k**, **l**). **m**, Schematic of CD38 regulation of mitochondria release/transfer hypothesis between astrocytes and neurons.

this extracellular population of astrocyte-derived mitochondria (Fig. 4d). At the same time, flow cytometry was used to quantify levels of MAP2 neuronal mitochondria (Extended Data Fig. 9). Brains treated with *Cd38* siRNA showed a significant reduction in neuronal mitochondria (Fig. 4e), suggesting that interfering with CD38 signalling may have suppressed endogenous astrocyte-to-neuron mitochondrial transfer. These effects were accompanied by a reduction in peri-infarct GAP43 (a surrogate marker of neuroplasticity; Fig. 4f, g) as well as worsened neurological outcomes (Fig. 4h, i). Furthermore, CD38 suppression significantly decreased oxygen consumption measurements in CSF-derived extracellular mitochondrial particles (Fig. 4j), and neurological outcomes seemed to be negatively correlated with these functional endpoints (Fig. 4k, l), suggesting that CSF mitochondrial function may be a potential biomarker of neuroglial signalling after stroke.

Taken together, these findings suggest that astrocytes may release extracellular mitochondrial particles via CD38-mediated mechanisms that enter neurons after stroke (Fig. 4m). However, there are a few caveats and the detailed mechanisms and generalizability of these proof-of-concept findings should warrant further investigation. First, the dynamics of extracellular mitochondria release and entry into neurons as well as quantitative thresholds for functional benefit remain to be fully defined (Extended Data Fig. 10a–i). A second caveat relates to mitochondrial entry mechanisms. In neurons, endocytosis may be regulated by dynamin/clathrin[20] or integrin pathways[21]. In our models, integrin-mediated Src/Syk signalling may be involved (Extended Data Fig. 10j–m). How integrin-mediated mitochondrial transfer is modulated under different disease conditions requires further study. Third, CD38 is also expressed in immune cells. In this study, CD38 suppression by siRNA *in vivo* did not appear to affect T cells or microglia/macrophages, but the balance between potentially beneficial CD38 signals in astrocytes versus deleterious CD38 signals in immune cells should be carefully considered. A fourth caveat is whether other glial cells may participate. Microglia, oligodendrocytes and pericytes are activated after stroke[22,23], so their potential roles in mitochondrial exchange warrant further investigation. Finally, astrocytes can produce many factors for protecting and restoring neurons, including tPA, high-mobility group box 1 (HMGB1), extracellular microvesicles containing VEGF and FGF-2, and various microRNAs[24–27]. How mitochondrial particles may interact with these other extracellular signals should be explored.

Non-cell autonomous signalling is vital for central nervous system recovery after injury or disease[28,29]. In the context of cerebral ischaemia, the present study suggests that astrocytes may release extracellular mitochondrial particles that enter neurons to support cell viability and recovery after stroke.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 January; accepted 14 June 2016.

1. Davis, C. H. *et al.* Transcellular degradation of axonal mitochondria. *Proc. Natl Acad. Sci. USA* **111,** 9633–9638 (2014).
2. Iadecola, C. & Nedergaard, M. Glial regulation of the cerebral microvasculature. *Nat. Neurosci.* **10,** 1369–1376 (2007).
3. Attwell, D. *et al.* Glial and neuronal control of brain blood flow. *Nature* **468,** 232–243 (2010).

4. Khakh, B. S. & Sofroniew, M. V. Diversity of astrocyte functions and phenotypes in neural circuits. *Nat. Neurosci.* **18,** 942–952 (2015).
5. Rosenberg, P. A. & Aizenman, E. Hundred-fold increase in neuronal vulnerability to glutamate toxicity in astrocyte-poor cultures of rat cerebral cortex. *Neurosci. Lett.* **103,** 162–168 (1989).
6. Wang, X. F. & Cynader, M. S. Pyruvate released by astrocytes protects neurons from copper-catalyzed cysteine neurotoxicity. *J. Neurosci.* **21,** 3322–3331 (2001).
7. Ouyang, Y. B. *et al.* Astrocyte-enriched miR-29a targets PUMA and reduces neuronal vulnerability to forebrain ischemia. *Glia* **61,** 1784–1794 (2013).
8. Nagai, M. *et al.* Astrocytes expressing ALS-linked mutated SOD1 release factors selectively toxic to motor neurons. *Nat. Neurosci.* **10,** 615–622 (2007).
9. Haidet-Phillips, A. M. *et al.* Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. *Nat. Biotechnol.* **29,** 824–828 (2011).
10. Voloboueva, L. A., Suh, S. W., Swanson, R. A. & Giffard, R. G. Inhibition of mitochondrial function in astrocytes: implications for neuroprotection. *J. Neurochem.* **102,** 1383–1394 (2007).
11. Anne Stetler, R., Leak, R. K., Gao, Y. & Chen, J. The dynamics of the mitochondrial organelle as a potential therapeutic target. *J. Cereb. Blood Flow Metab.* **33,** 22–32 (2013).
12. Falchi, A. M. *et al.* Astrocytes shed large membrane vesicles that contain mitochondria, lipid droplets and ATP. *Histochem. Cell Biol.* **139,** 221–231 (2013).
13. Islam, M. N. *et al.* Mitochondrial transfer from bone-marrow-derived stromal cells to pulmonary alveoli protects against acute lung injury. *Nat. Med.* **18,** 759–765 (2012).
14. Aksoy, P., White, T. A., Thompson, M. & Chini, E. N. Regulation of intracellular levels of NAD: a novel role for CD38. *Biochem. Biophys. Res. Commun.* **345,** 1386–1392 (2006).
15. Guse, A. H. & Lee, H. C. NAADP: a universal Ca^{2+} trigger. *Sci. Signal.* **1,** re10 (2008).
16. Bruzzone, S. *et al.* Glutamate-mediated overexpression of CD38 in astrocytes cultured with neurones. *J. Neurochem.* **89,** 264–272 (2004).
17. Levy, A. *et al.* CD38 facilitates recovery from traumatic brain injury. *J. Neurotrauma* **26,** 1521–1533 (2009).
18. Higashida, H. *et al.* Social memory, amnesia, and autism: brain oxytocin secretion is regulated by NAD+ metabolites and single nucleotide polymorphisms of CD38. *Neurochem. Int.* **61,** 828–838 (2012).
19. Choe, C. U. *et al.* CD38 exacerbates focal cytokine production, postischemic inflammation and brain injury after focal cerebral ischemia. *PLoS One* **6,** e19046 (2011).
20. Frühbeis, C. *et al.* Neurotransmitter-triggered transfer of exosomes mediates oligodendrocyte-neuron communication. *PLoS Biol.* **11,** e1001604 (2013).
21. Bowen, S. *et al.* The phagocytic capacity of neurones. *Eur. J. Neurosci.* **25,** 2947–2955 (2007).
22. Winkler, E. A., Bell, R. D. & Zlokovic, B. V. Central nervous system pericytes in health and disease. *Nat. Neurosci.* **14,** 1398–1405 (2011).
23. Hu, X. *et al.* Microglial and macrophage polarization—new prospects for brain repair. *Nat. Rev. Neurol.* **11,** 56–64 (2015).
24. Xin, H. *et al.* Increasing tPA activity in astrocytes induced by multipotent mesenchymal stromal cells facilitate neurite outgrowth after stroke in the mouse. *PLoS One* **5,** e9027 (2010).
25. Hayakawa, K., Pham, L. D., Katusic, Z. S., Arai, K. & Lo, E. H. Astrocytic high-mobility group box 1 promotes endothelial progenitor cell-mediated neurovascular remodeling during stroke recovery. *Proc. Natl Acad. Sci. USA* **109,** 7505–7510 (2012).
26. Proia, P. *et al.* Astrocytes shed extracellular vesicles that contain fibroblast growth factor-2 and vascular endothelial growth factor. *Int. J. Mol. Med.* **21,** 63–67 (2008).
27. Li, Y., Liu, Z., Xin, H. & Chopp, M. The role of astrocytes in mediating exogenous cell-based restorative therapy for stroke. *Glia* **62,** 1–16 (2014).
28. Lo, E. H. Degeneration and repair in central nervous system disease. *Nat. Med.* **16,** 1205–1209 (2010).
29. Xing, C. & Lo, E. H. Help-me signaling: Non-cell autonomous mechanisms of neuroprotection and neurorecovery. *Prog. Neurobiol.* http://dx.doi.org/10.1016/j.pneurobio.2016.04.004 (2016).

Acknowledgements This work was supported in part by grants from the National Institutes of Health (NIH), the Rappaport Foundation, and the China National Natural Science Foundation Award For Distinguished Young Scholars. Electron microscopy was performed in the Center for Systems Biology. Cytometric assessments were supported by the Department of Pathology Flow and Image Cytometry Core. The authors thank J. Felton and J. Zwicker for assistance with qNano analysis.

Author Contributions K.H. contributed to manuscript preparation, hypothesis generation, experimental design/analysis and conducted experiments. E.E., X.W., Y.T., Y.L. and C.X. conducted experiments and helped with data analysis. X.J. and E.H.L. contributed to manuscript preparation, hypothesis generation and experimental design.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.H. (khayakawa1@mgh.harvard.edu), X.J. (jixunming@vip.163.com) or E.H.L. (Lo@helix.mgh.harvard.edu).

METHODS

No statistical methods were used to predetermine sample size.

Reagents. BAPTA-AM (A1076), cADPR (C7344), β-nicotinamide adenine dinucleotide sodium salt (N0632) and dynasore hydrate (D7693) were purchased from Sigma, and RGDS peptide (3498) and MNS (2877/50) were from R&D systems.

Mouse focal cerebral ischaemia model. All experiments were performed following an institutionally approved protocol in accordance with National Institutes of Health guidelines and with the United States Public Health Service's Policy on Human Care and Use of Laboratory Animals. Our methods also included randomization, blinding and statistical criteria consistent with ARRIVE guidelines (Animals in Research: Reporting *In vivo* Experiments). In brief, male C57BL/6J (12–14 weeks) or FVB/N-Tg (GFAPGFP)14Mes/J mice are anaesthetized with 5% to 1% isoflurane, and rectal temperatures and cerebral blood flow are monitored. After midline skin incision, 7-0 nylon monofilament coated with silicon resin was introduced through a small incision into the common carotid artery. Adequate cerebral ischaemia was assessed by Laser Doppler flowmetry and by examining forelimb flexion after the mice recovered from anaesthesia. The mice were reanaesthetized, and reperfusion was established by withdrawal of the filament. Functional outcome after stroke was assessed by neurological severity scores and the foot-fault test[25].

Primary neuron cultures. Primary neuron cultures were prepared from cerebral cortices of embryonic day (E)17 Sprague–Dawley rat embryos or E17 FVB/N-Tg (GFAPGFP)14Mes/J mouse embryos. In brief, cortices were dissected and dissociated using papain dissociation system (Worthington Biochemical Corporation, LK003150). Cells were spread on plates coated with poly-D-lysine (Sigma, P7886) and cultured in DMEM (NBM, Life Technology, 11965-084) containing 25 mM glucose, 4 mM glutamine, 1 mM sodium pyruvate, and 5% FBS at a density of 2×10^5 cells ml^{-1} (1 ml for 12-well format, 0.5 ml for 24-well format). At 24 h after seeding, the medium was changed to Neurobasal medium (Invitrogen, 21103-049) supplemented with B-27 (Invitrogen, 17504044) and 0.5 mM glutamine. Cells were cultured at 37 °C in a humidified chamber of 95% air and 5% CO_2. Cultures were used for experiments from 7 to 10 days after seeding.

Primary astrocyte cultures. Primary astrocyte cultures were prepared from cerebral cortices of 2-day-old neonatal Sprague–Dawley rats or E17 C57BL/6J mice. In brief, dissociated cortical cells were suspended in DMEM (Life Technology, 11965-084) containing 25 mM glucose, 4 mM glutamine, 1 mM sodium pyruvate, and 10% FBS and plated on uncoated 25 cm^2 flasks at a density of 6×10^5 cells cm^{-2}. Monolayers of type 1 astrocytes were obtained 12–14 days after plating. Non-astrocytic cells such as microglia and neurons were detached from the flasks by shaking and removed by changing the medium. Astrocytes were dissociated by trypsinization and then reseeded on uncoated T75 flasks. After the cells reached 70–80% confluence, cultures were switched to Neurobasal medium containing 1% penicillin/streptomycin or DMEM containing 1% penicillin/streptomycin, and astrocyte-conditioned media (ACM) were collected 24 h later. Collected ACM was treated by spin cell debris down with centrifuging at 2,000*g* for 10 min or by filtrating through a 1.2-μm syringe filter for further experiments.

OGD and reoxygenation. OGD experiments were performed using a specialized, humidified chamber (Heidolph, incubator 1000, Brinkmann Instruments) kept at 37 °C, which contained an anaerobic gas mixture (90% N_2, 5% H_2, and 5% CO_2). To initiate OGD, culture medium was replaced with deoxygenated, glucose-free DMEM (Life Technology, 11966-025). After 2 h challenge, cultures were removed from the anaerobic chamber, and the OGD solution in the cultures was replaced with maintenance medium. Cells were then allowed to recover for 18 h (for neurotoxicity assay) and 72 h (for siRNA/astrocyte coculture) in a regular incubator.

Cell viability assays. Neuronal injury was measured by standard cell cytotoxicity assays such as lactate dehydrogenase (LDH) using the Cytotoxicity Detection Kit (Roche Applied Science, 11644793001) and/or Cell Counting Kit 8 cytotoxicity assay (DOJINDO, CK04-13). For the LDH assay, 100% cell death was induced with 0.5% Triton X-100 in sister culture. The relative assessments of neuronal injury were normalized by comparison with 100% cell death (LDH assay) or with control cell as 100% cell survival (CCK8).

Determination of CD38/ADPR-cyclase activity. ADPR cyclase activity was determined fluorometrically using nicotinamide guanine dinucleotide (NGD^+) (Sigma, N5131) as a substrate as described previously[30,31]. Astrocytes or neurons were incubated with 200 μM NGD^+, and the production of cGDPR was determined at excitation/emission wavelengths of Ex 300 nm/Em 410 nm with a microplate reader.

ATP measurement. Intracellular or extracellular ATP was determined by CellTiter-Glo luminescence kit (Promega, G7570), which can perform cell lysis and generate a luminescent signal proportional to the amount of ATP present. In brief, opaque-walled 96-well plates with culture media (50 μl) or cell lysate (50 μl) were prepared. CellTiter-Glo luminescence test solution (50 μl) was added and incubated for 30 min at room temperature. Luminescent signal was determined by luminescence microplate reader.

Liposomal ATP treatment. Liposomal ATP was obtained from Encapsula NanoScience. In brief, we used lyophilized proliposomes compose of 7:3 molar ratio of L-α-phosphatidylcholine: L-α-phosphatidylserine containing ATP which forms 100 nm liposomal ATP upon hydration. ATP-loaded liposome (1–1,000 nM) was co-incubated with neurons following OGD, and cell viability was analysed after 18 h reoxygenation.

Mitochondria membrane potential measurement. To monitor mitochondrial health, JC-1 dye (Invitrogen, T-3168) was used to assess mitochondrial membrane potential. Rat cortical astrocytes or media were incubated with JC1 (5 μM or 1 μM) for 30 min at 37 °C. JC1 dye exhibits potential-dependent accumulation in mitochondria, indicated by fluorescence emission shift from green (Ex 485 nm/Em 516 nm) to red (Ex 579 nm/Em 599 nm). Mitochondria membrane potential was determined by the fluorescent ratio with a fluorescent microplate reader.

Oxygen consumption analysis. Real-time oxygen consumption in astrocytic particles or in CSF samples was measured by the Mito-ID Extracellular O2 sensor kit (Enzo Life Science, ENZ-51045) according to the instruction provided Enzo Life Science. In brief, astrocytes (70–80% confluent cells per well per 100 μl) or particle fractions (100 μl; 25-fold concentrated astrocytic conditioned media) were prepared in non-coated regular 96 wells, and O_2 sensor probe (10 μl) was added into each well. Each CSF sample (8–20 μl) was collected from cisterna magna at day 7 after focal cerebral ischaemia. After centrifugation at 2,000*g* for 10 min, 6 μl CSF was diluted in 54 μl PBS, and 6 μl O_2 sensor probe was added into each well. After covering with 100 μl (50 μl for CSF sample) of Mito-ID HS Oil, the plates were read with filter combination of 340 nm for excitation and 642 nm of emission at 30 °C.

Electron microscopy analysis. Rat cortical astrocytes or pellets from astrocyte-conditioned media were fixed in 2.0% glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.4 (Electron Microscopy Sciences) for 1 h at room temperature on a rocker. They were rinsed in cacodylate buffer, gently scraped and pelleted and post-fixed in 1.0% osmium tetroxide in cacodylate buffer for 1 h on ice. They were rinsed in buffer and stabilized with a small amount of 2% agarose in PBS to hold them together. They were then dehydrated through a graded series of ethanol to 100%, followed by propylene oxide, 100%. They were infiltrated with Epon resin (Ted Pella) in a 1:1 solution of Epon:propylene oxide overnight on a rocker at room temperature. The next day they were placed in fresh Epon for several hours and then embedded in Epon overnight at 60 °C. Thin sections were cut on a Leica EM UC7 ultramicrotome, collected on formvar-coated grids, stained with uranyl acetate and lead citrate and examined in a JEOL JEM 1011 transmission electron microscope at 80 kV. Images were collected using an AMT digital imaging system (Advanced Microscopy Techniques). These methods are similar to previous descriptions of extracellular particle and mitochondria detection in astrocyte cultures[12].

FACS analysis. Standard FACS analysis was performed by BD LSR II or BD Fortessa as described previously[25,32,33]. ACM was collected from rat cortical astrocytes labelled with MitoTracker Red CMXRos followed by filtrating through a 1.2-μm syringe filter. The supernatant was used to sort labelled mitochondria fraction by FACSAriaII cell sorter configured with 561 nm air cooled laser. Brain cells were collected from peri-infarct cortex after stroke. In brief, tissues are gently minced and then digested at 37 °C for 30 min with an enzyme cocktail (collagenase type I, DNase I, Sigma Aldrich). CSF samples were prepared for further staining after centrifugation at 2,000*g* for 10 min. FACS analysis was performed using an unstained or phenotype control for determining appropriate gates, voltages, and compensations required in multivariate flow cytometry.

Measurement of particle size. Particle size following extracellular mitochondria isolation by FACS was determined by qNano (iZON). Nanopore-based detection allows particle-by-particle assessment of complex mixtures. Optimization of pore size to particle size, by adjusting the stretch of the pore, allows highly accurate measurement[34]. Particles containing mitochondria were sorted using FACS analysis, then particle sizes were quantified by NP400 and using CPC400 calibration particle.

Western blot analysis. Western blot was performed as previously reported[25]. Pro-PREP Protein Extraction Solution (iNtRON Biotechnology, 17081) or Qproteome FFPE Tissue kit (Qiagen, 37623) was used to collect samples. Each sample was loaded onto 4–20% Tris-glycine gels. After electrophoresis and transferring to nitrocellulose membranes, the membranes were blocked in Tris-buffered saline containing 0.1% Tween 20 and 0.2% I-block (Tropix, T2015) for 90 min at room temperature. Membranes were then incubated overnight at 4 °C with the following primary antibodies, anti-β-actin (1:1,000, Sigma-Aldrich, A5441), anti-GFAP antibody (1:1,000, BD Biosciences, 556328), anti-MAP2 antibody (1:500, Abcam, ab11267 or ab32454), anti-CD38 antibody (1:500, Santacruz, sc-7049), anti-TOM40 (1:200, Santacruz, sc-11414), anti-phosphorylated AKT (1:500, Cell Signaling, 9271), anti-BCL-XL (1:500, Cell Signaling, 2764), anti-active caspase 3

(1:200, Abcam, ab32042), anti-AIF (1:500, Abcam, ab32516), and anti-GAP43 (1:500, Santacruz, sc-17790). After incubation with peroxidase-conjugated secondary antibodies, visualization was enhanced by chemiluminescence (GE Healthcare, NA931 (anti-mouse), NA934 (anti-rabbit), or NA935 (anti-rat)). Optical density was assessed using the NIH Image analysis software.

Immunocytochemistry and immunohistochemistry. Immunocytochemistry and immunohistochemistry were performed as described previously[35,36]. After staining with primary antibody and fluorescent-tagged secondary antibody, nuclei were counterstained with or without 4,6-diamidino-2-phenylindole (DAPI), and coverslips were placed. Immunostaining images or time lapse images were obtained with a fluorescence microscope (Nikon ECLIPSE Ti-S) interfaced with a digital charge-coupled device camera and an image analysis system or confocal microscope analysis using Carl Zeiss Laser Scanning Confocal Microscope Pascal 5 LSM and Pascal 5 LSM software Version 3.2. Dendrite elongation was assessed following MAP2 staining followed by NeuriteQuant analysis[37].

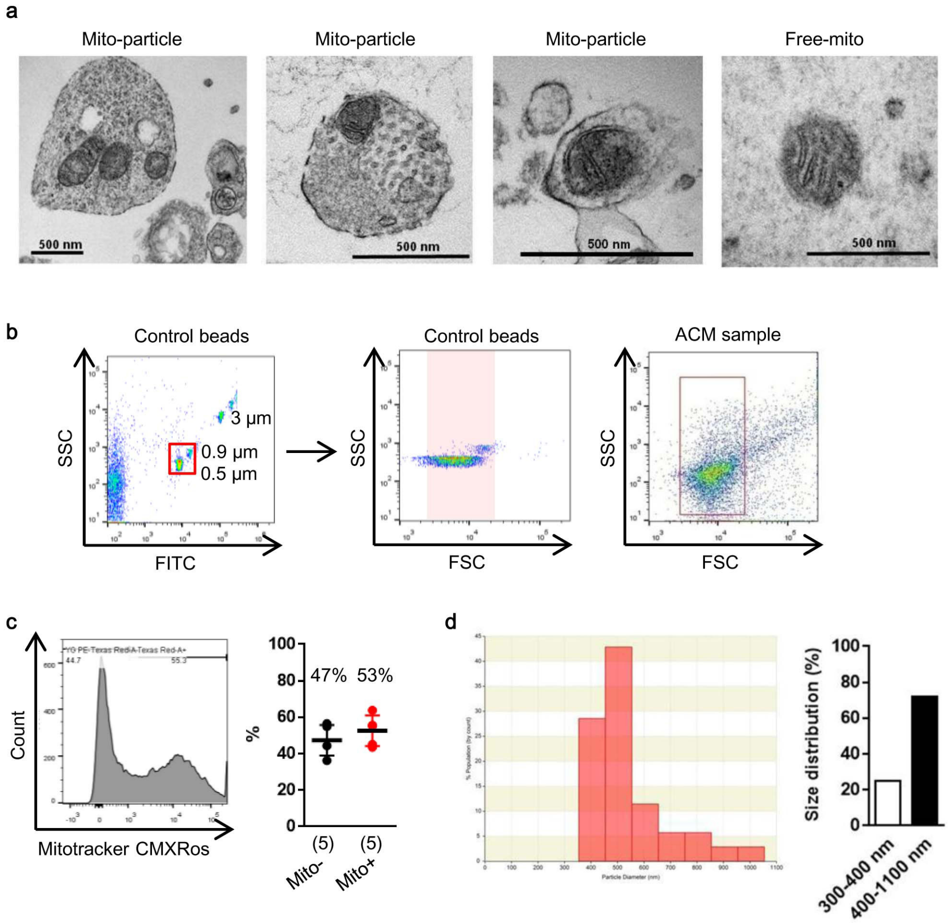
CRISPR activation plasmid transfection. Control CRISPR activation plasmid (sc-437275), and rat CD38 CRISPR activation plasmid (sc-437321-ACT) were obtained from Santa Cruz Biotechnology. Transfection was performed according to the transfection protocol for cell cultures from Santa Cruz Biotechnology. In brief, 1 ml of plasmid transfection reagent mixture (transfection reagent, sc-395739, transfection medium, sc-108062) was co-incubated with astrocytes for 24 h in a 5% CO_2 incubator at 37 °C, and then CD38 cyclase activity was assessed to confirm efficiency of transfection.

siRNA experiments. Control siRNA and *Cd38* siRNA were obtained from Santa Cruz Biotechnology. Control siRNA (sc-37007) consists of a scrambled sequence that will not lead to the specific degradation of any known cellular mRNA. Mouse *Cd38* siRNA (sc-37246) or rat *Cd38* siRNA (sc-270394) is each pool of 3 target-specific 19–25-nucleotide siRNAs designed to knock down gene expression. The sequences for mouse *Cd38* siRNAs are: 5′-GUGUACUACCAACAUUCAA-3′, 5′-GUGUGUCUUUAGUAGGUAU-3′, 5′-CCAGUUUGUGAUUGUUGA-3′. Rat *Cd38* siRNAs are: sequence 1: 5′-CUCAAACCAUACCAUGUAA-3′, sequence 2: 5′-GGAAGAGCUUCCCAAUACA-3′, sequence 3: 5′-GUGUUAU CGUCUAGCAAUA-3′.

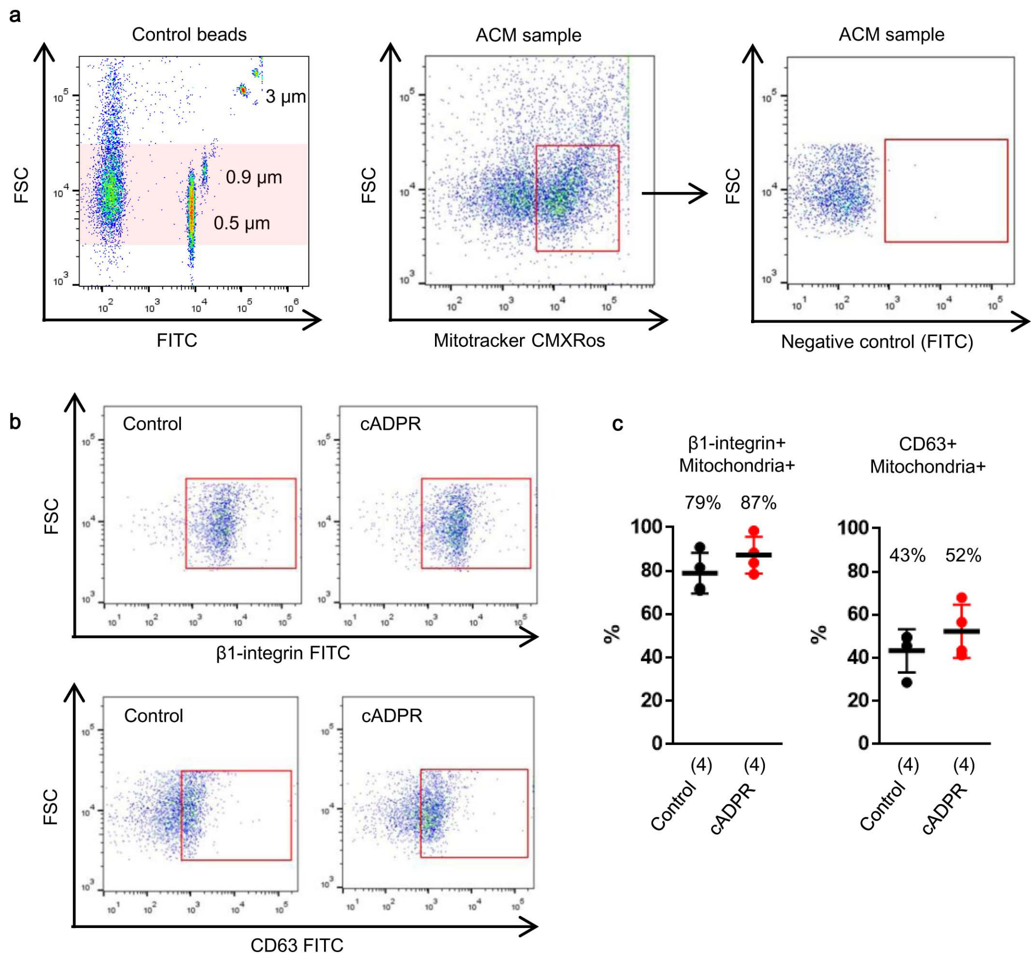
siRNA was prepared according to the transfection protocol for cell cultures from Santa Cruz Biotechnology. In brief, 1 ml of siRNA transfection reagent mixture (transfection reagent, sc-29528, transfection medium, sc-36868) was co-incubated with astrocytes for 6 h in a 5% CO_2 incubator at 37 °C, and then the same amount of DMEM 20% FBS was added. An additional incubation was performed for 18 h.

Statistical analysis. Results were expressed as mean ± s.e.m. When only two groups were compared, an unpaired *t*-test was used. Multiple comparisons were evaluated by one-way ANOVA followed by Tukey's test, two-way ANOVA or repeated-measure two-way ANOVA followed by Bonferroni test. Correlation was evaluated by Spearman's rank correlation. $P < 0.05$ was considered statistically significant. Statistical analyses were performed using GraphPad Prism 6.

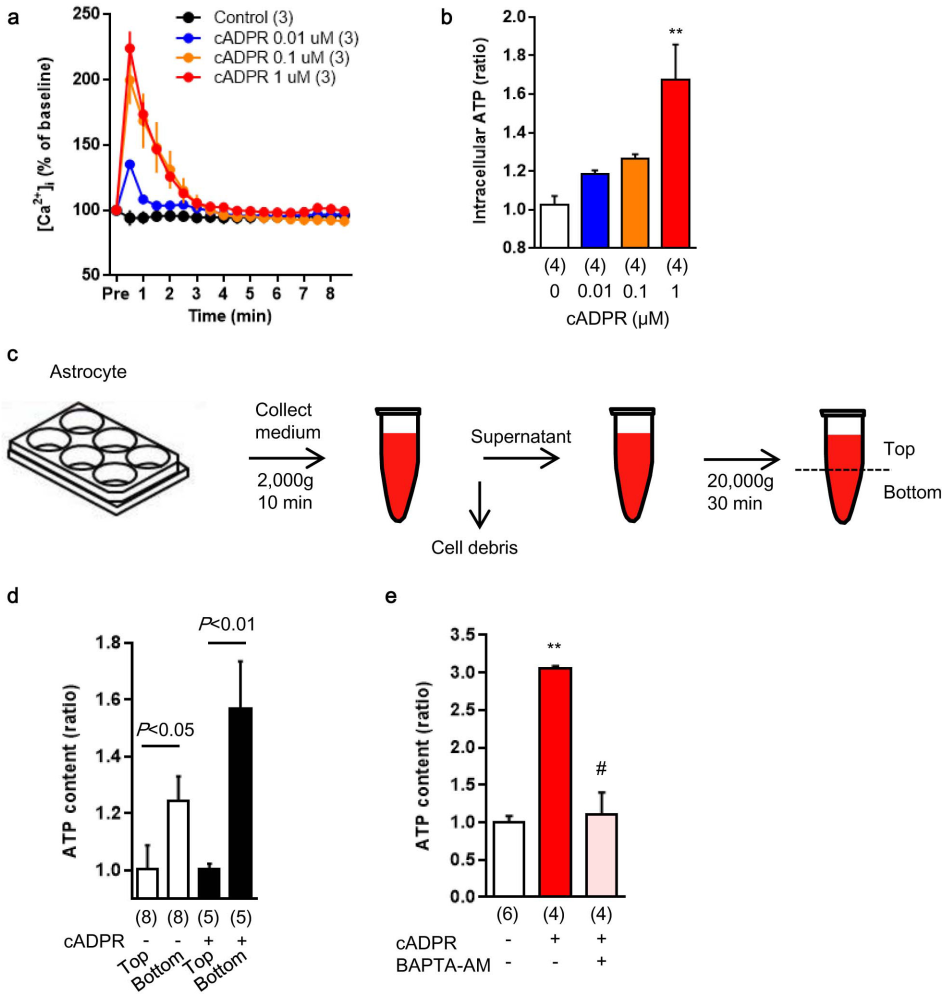
30. Rah, S. Y., Park, K. H., Han, M. K., Im, M. J. & Kim, U. H. Activation of CD38 by interleukin-8 signaling regulates intracellular Ca^{2+} level and motility of lymphokine-activated killer cells. *J. Biol. Chem.* **280,** 2888–2895 (2005).
31. Graeff, R. M., Walseth, T. F., Fryxell, K., Branton, W. D. & Lee, H. C. Enzymatic synthesis and characterizations of cyclic GDP-ribose. A procedure for distinguishing enzymes with ADP-ribosyl cyclase activity. *J. Biol. Chem.* **269,** 30260–30267 (1994).
32. Bi, B. *et al.* Cortical glial fibrillary acidic protein-positive cells generate neurons after perinatal hypoxic injury. *J. Neurosci.* **31,** 9205–9221 (2011).
33. Cruz, F. C. *et al.* New technologies for examining the role of neuronal ensembles in drug addiction and fear. *Nat. Rev. Neurosci.* **14,** 743–754 (2013).
34. Vogel, R. *et al.* Quantitative sizing of nano/microparticles with a tunable elastomeric pore sensor. *Anal. Chem.* **83,** 3499–3506 (2011).
35. Hayakawa, K. *et al.* Inhibition of reactive astrocytes with fluorocitrate retards neurovascular remodeling and recovery after focal cerebral ischemia in mice. *J. Cereb. Blood Flow Metab.* **30,** 871–882 (2010).
36. Hayakawa, K., Arai, K. & Lo, E. H. Role of ERK map kinase and CRM1 in IL-1β-stimulated release of HMGB1 from cortical astrocytes. *Glia* **58,** 1007–1015 (2010).
37. Dehmelt, L., Poplawski, G., Hwang, E. & Halpain, S. NeuriteQuant: an open source toolkit for high content screens of neuronal morphogenesis. *BMC Neurosci.* **12,** 100 (2011).



Extended Data Figure 1 | Astrocytic mitochondria particle detection. **a**, Electron microscopic analysis demonstrated that mitochondria were detected within extracellular astrocyte-derived particles. Free mitochondria were also found in ACM. **b**, In FACS analysis, control beads were used to gate populations ranging in size from 500 to 900 nm. **c**, In ACM, approximately 53% of particles in the size range were positive for functional mitochondria ($n = 2$ biological replicates, $n = 5$ independent experiments). **d**, After FACS analysis to isolate the extracellular mitochondria fraction from ACM, particle size was measured by qNano analysis. Consistent with electron microscopy analysis, a range of size distributions was observed (~25%: 300–400 nm, ~75%: 400–1,100 nm). Data are mean ± s.e.m.

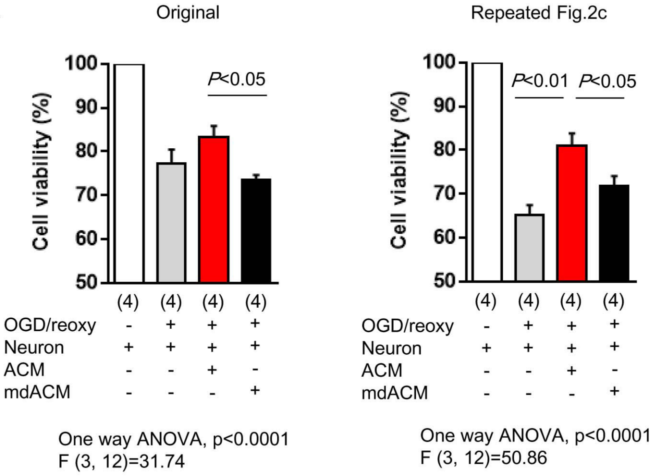


Extended Data Figure 2 | Characteristics of astrocytic mitochondrial particles in FACS analysis. a, Mitochondrial particles were identified by FACS. **b**, Of these mitochondrial particles, FACS analysis showed that approximately 79% and 43% of particles express β1-integrin and CD63, respectively ($n = 2$ biological replicates, $n = 4$ independent experiments). cADPR (1 µM) did not appear to affect these distributions ($n = 2$ biological replicates, $n = 4$ independent experiments). Data are mean ± s.e.m.

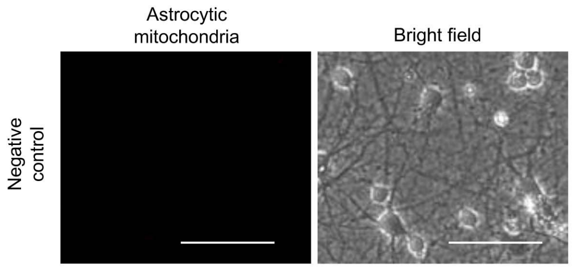


Extended Data Figure 3 | Production of astrocytic mitochondrial particles in a Ca^{2+}-dependent mechanism. a, The known CD38 downstream signal, cADPR, increased intracellular calcium levels shown by Fluo-4 intensity in a concentration-dependent manner ($n = 3$ independent experiments). **b**, Intracellular ATP levels in astrocytes were upregulated by cADPR stimulation ($n = 4$ independent experiments). $^{**}P < 0.01$ versus 0 μM cADPR. **c**, To measure ATP levels in extracellular particles, ACM was collected and large debris was excluded by centrifugation and filtration through a 1.2-μm filter. After another centrifugation at 20,000*g* for 30 min, 100 μl from the top or bottom fractions were used for the ATP assay. **d**, The bottom fraction had a higher ATP content, and cADPR (1 μM) increased the ATP content in this fraction ($n = 2$ biological replicates, $n = 8$ or 5 independent experiments). **e**, cADPR-induced extracellular ATP levels within extracellular particles were diminished by the intracellular calcium blocker, BAPTA-AM ($n = 2$ biological replicates, $n = 6$ or 4 independent experiments). Data are mean ± s.e.m. *P* values are from a one-way ANOVA followed by Tukey's test.

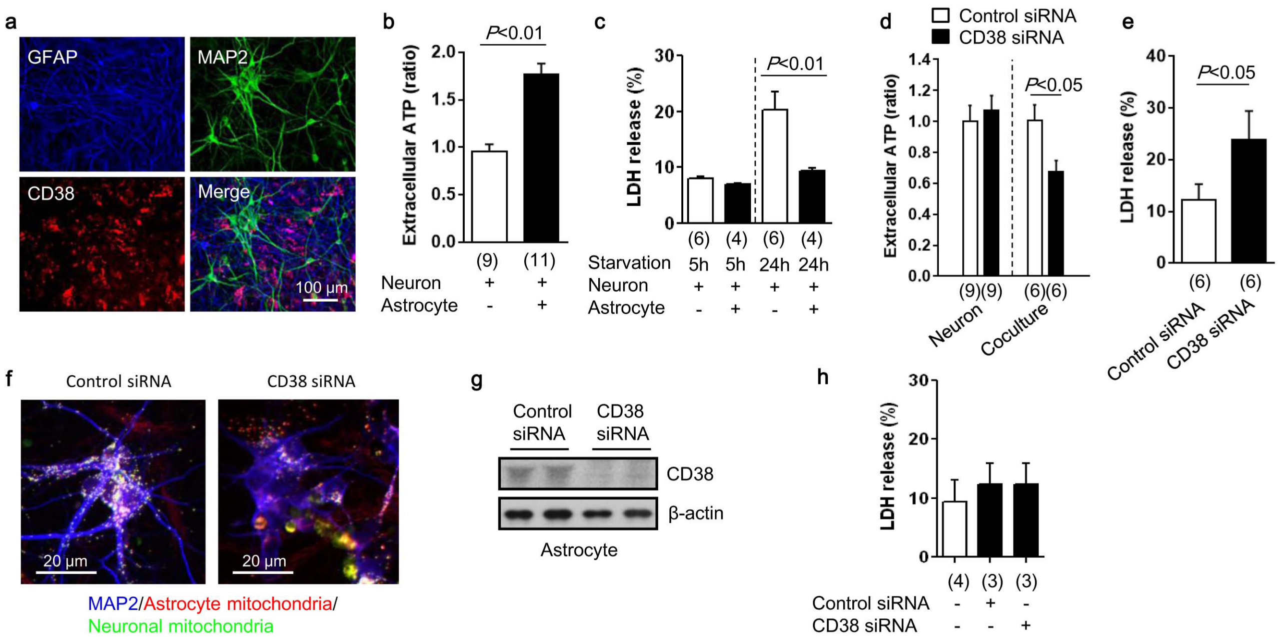
a



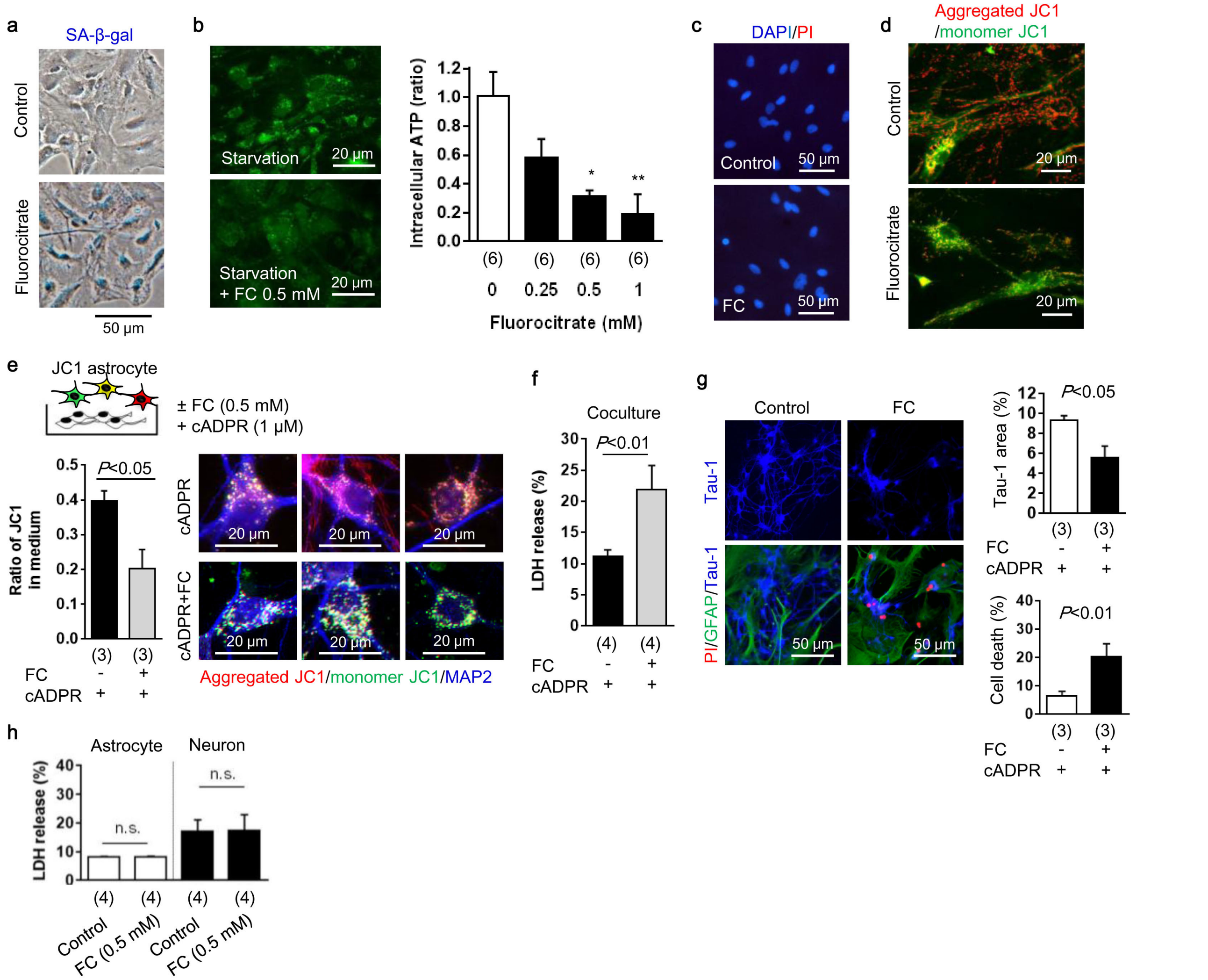
b



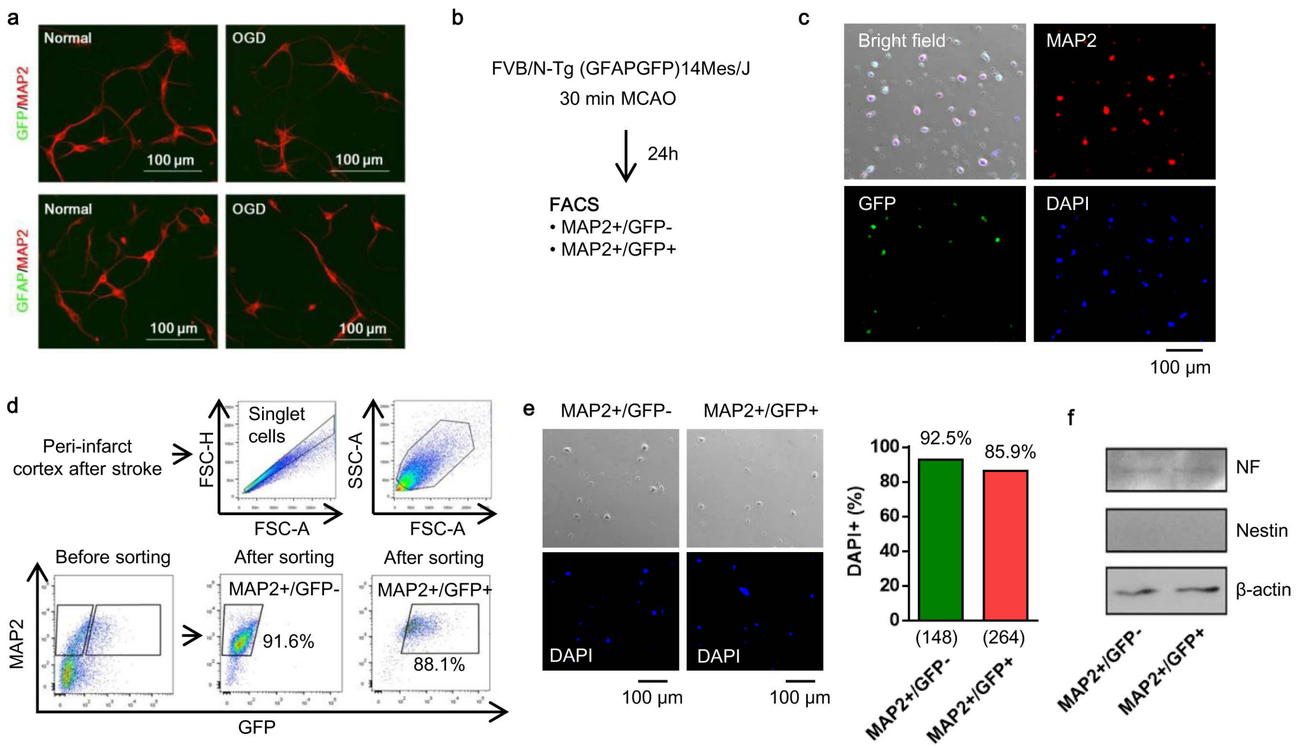
Extended Data Figure 4 | Summary of experiment in Fig. 2c. **a**, We repeated the experiment in Fig. 2c with $n = 4$ independent primary cultures per group. Similar results were obtained ($n = 2$ biological replicates, $n = 4$ independent experiments). The extracellular mdACM group was significantly different compared to the ACM group. Furthermore, in this repeated experiment, there was also statistical significance between controls (OGD-damaged neurons alone) versus those treated with mitochondria-containing astrocyte media (ACM), and there was no statistically significant worsening when comparing control versus mitochondria-depleted groups (mdACM). Taken together, these two separate experiments suggest a modest but statistically significant neuroprotection induced by astrocyte-derived mitochondria. Data are mean ± s.e.m. one-way ANOVA followed by Tukey's test. **b**, MitoTracker Red CMXRos (200 nM) was incubated without astrocytes to obtain no-cell-derived media (negative control). Media was collected and further incubated with neurons after OGD. After 24 h, there was no mitochondrial signal observed. Scale bars, 100 μm.



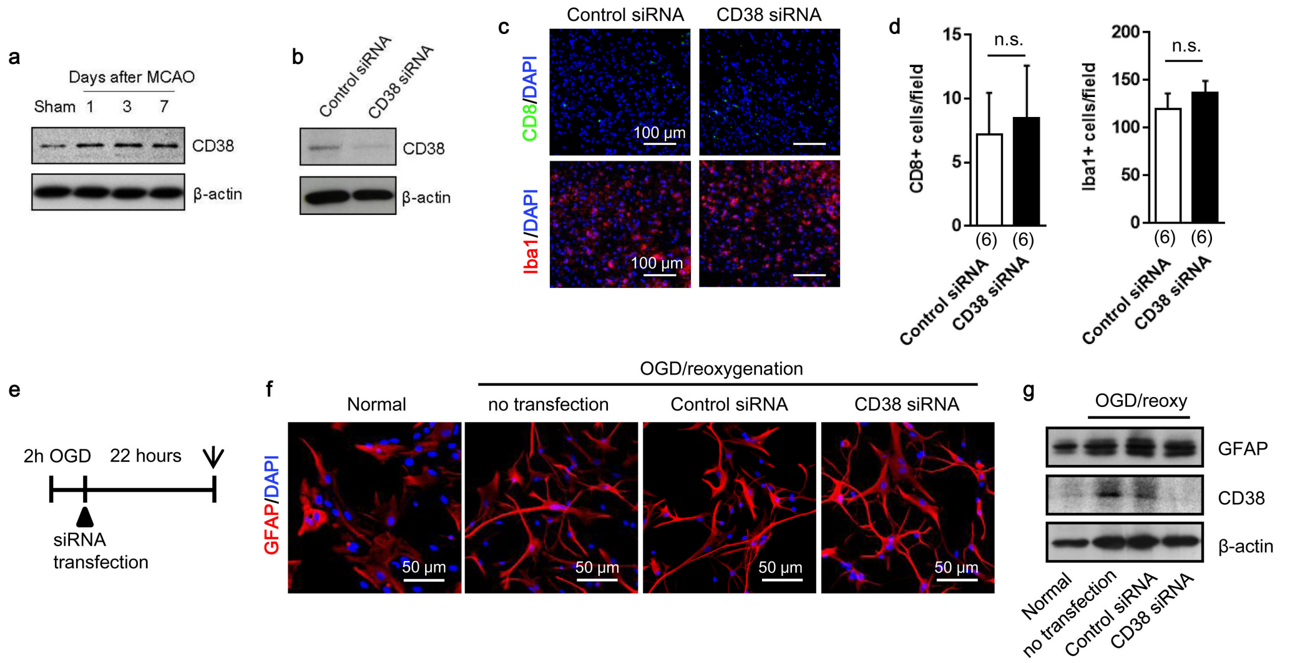
Extended Data Figure 5 | Role of astrocytic CD38 in mitochondria transfer during starvation *in vitro*. a, Immunocytochemistry in neuron–astrocyte co-cultures demonstrated that CD38 was primarily expressed within astrocytes. **b**, Extracellular ATP levels were higher in media collected from neurons co-cultured with astrocytes compared to media from neuron-alone cultures ($n = 3$ biological replicates, $n = 9$ or 11 independent experiments). **c**, After serum/glucose starvation, neurons were significantly damaged, as expected. But neurons co-cultured with astrocytes were protected ($n = 2$ biological replicates, $n = 6$ or 4 independent experiments). **d**, CD38 suppression by siRNA significantly decreased extracellular ATP levels in neuron–astrocyte co-cultures, but CD38 suppression did not affect extracellular ATP levels in neuron-alone cultures ($n = 3$ biological replicates, $n = 9$ or 6 independent experiments). **e**, Blockade of astrocytic CD38 by siRNA significantly increased LDH release (indicative of cell damage) in the co-cultures, suggesting that CD38 may be important for maintaining neuroglial homeostasis ($n = 2$ biological replicates, $n = 6$ independent experiments). **f**, Rat primary neurons were co-cultured with rat astrocytes. Immunocytochemistry showed that CD38 suppression by siRNA reduced astrocytic mitochondria (red) transfer into neurons compared to control. **g**, **h**, Western blot analysis indicated that CD38 suppression by siRNA can be successfully performed in astrocyte culture without affecting cell viability ($n = 2$ biological replicates, $n = 4$ or 3 independent experiments). Data are mean ± s.e.m. P values are from an unpaired t-test.



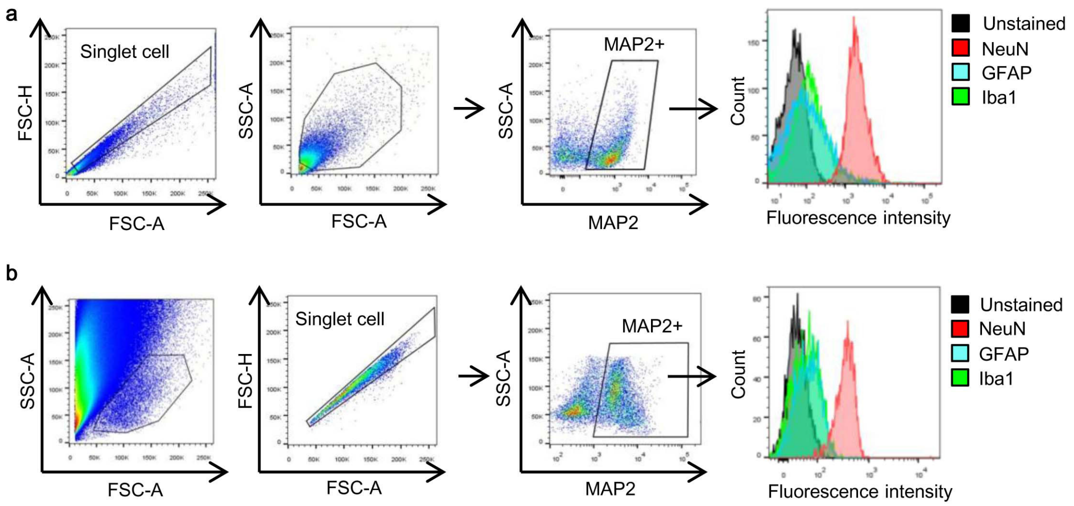
Extended Data Figure 6 | Metabolic inhibition in astrocyte causes neuronal cell death and prevents neurite outgrowth *in vitro*.
a, Astrocytic aconitase was inhibited by fluorocitrate (FC), which disrupted astrocyte metabolism that was accompanied by a senescence-associated β-galactosidase (SA-β-gal) signal. **b**, Intracellular ATP was decreased in these metabolically disrupted astrocytes ($n = 2$ biological replicates, $n = 6$ independent experiments). $*P < 0.05$, $**P < 0.01$ versus 0 mM fluorocitrate. **c**, Propidium iodide (PI) staining showed that fluorocitrate (0.5 mM) did not induce cell death in astrocytes. **d**, Metabolically disrupted astrocytes significantly decreased the mitochondrial membrane potential. Red: aggregated JC1; green: monomer JC1. Scale bar, 20 μm. **e**, Rat cortical neurons were co-cultured with JC1-labelled astrocytes. After 24 h co-culture, control astrocytes transferred mitochondria, which had a high-membrane potential (aggregated JC1), but metabolically disrupted astrocytes released and transferred dysfunctional mitochondria into neurons ($n = 3$ independent experiments). **f**, Metabolically disrupted astrocytes could not support neural viability under starvation in the co-culture ($n = 4$ independent experiments). **g**, Co-culture between astrocytes and neurons was conducted for 48 h to test neurite outgrowth. Immunocytochemistry showed that metabolically disrupted astrocytes prevented neurite outgrowth and increased neuronal cell death ($n = 3$ independent experiments). **h**, The LDH assay indicated that fluorocitrate (0.5 mM) did not affect cell viability in either rat cortical astrocytes ($n = 4$ independent experiments) or rat cortical neurons ($n = 4$ independent experiments). Data are mean ± s.e.m. P values are from a one-way ANOVA followed by Tukey's test (**b**) or an un unpaired t-test (**e**–**g**).



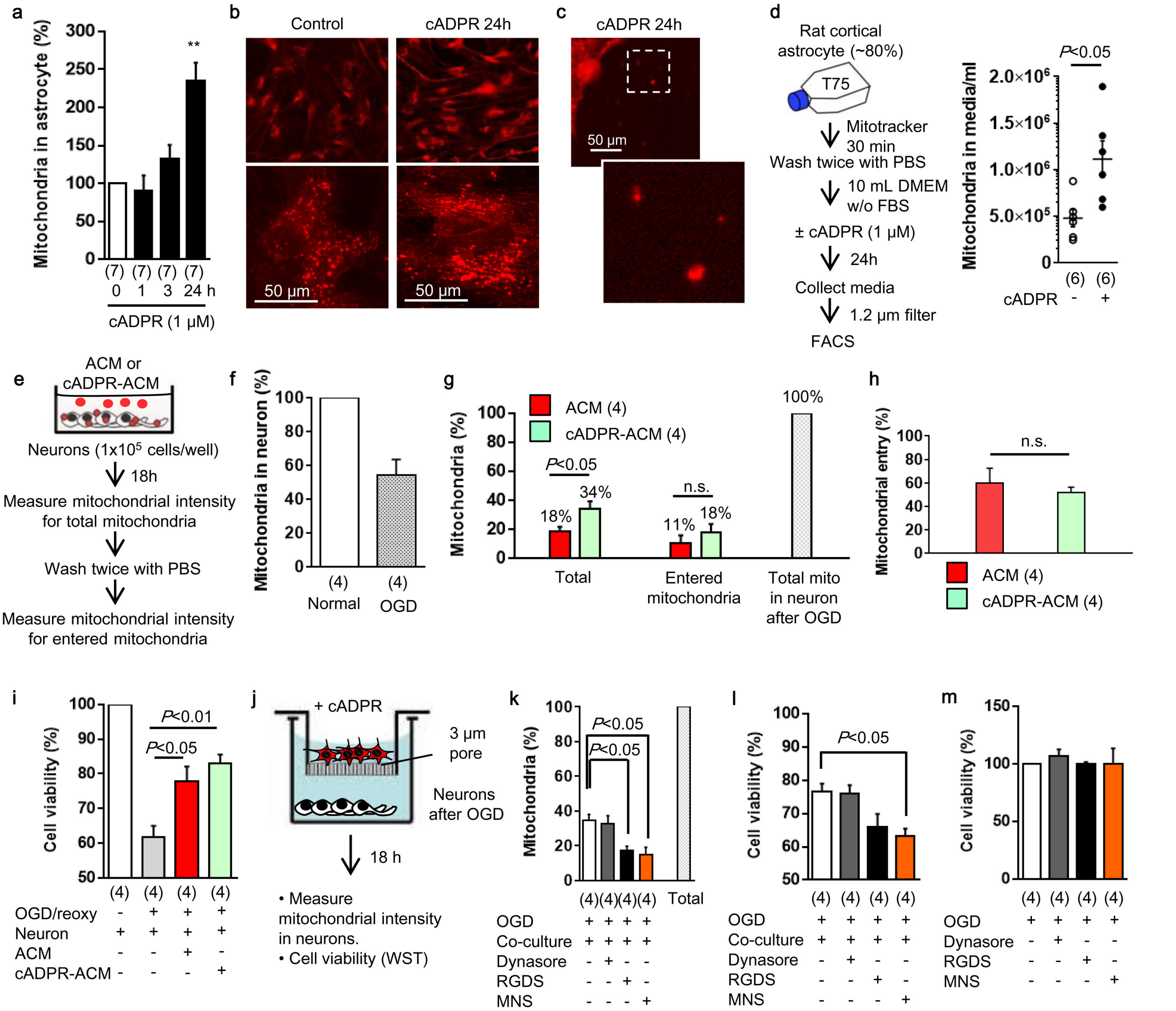
Extended Data Figure 7 | FACS analysis using E17 FVB/N-Tg (GFAPGFP)14Mes/J transgenic mice. a, Cortical neurons were isolated from E17 FVB/N-Tg (GFAPGFP)14Mes/J transgenic mice. Immunocytochemistry showed that cultured neurons did not express either GFP or GFAP protein after OGD, suggesting that stroke-like stress may not lead to 'leakiness' in this astrocyte-specific GFP mouse. **b**, Brain cell suspension was prepared from FVB/N-Tg (GFAPGFP)14Mes/J mice subjected to transient ischaemia, then FACS analysis was performed. **c**, Representative image before cell sorting. **d**, Purity after cell sorting. **e**, The MAP2^{+} GFP^{-} or MAP2^{+} GFP^{+} populations were 92.5% or 85.9% positive for DAPI, respectively. **f**, Western blot analysis demonstrated that both GFP-positive and -negative neurons expressed mature neuron marker (neurofilament) but not neuronal stem-cell marker (nestin). These data exclude the possibility that GFAP-positive cells included subsets of neuronal precursor cells that are known to also express GFAP.



Extended Data Figure 8 | Effects of CD38 suppression by siRNA *in vivo* and *in vitro*. **a**, Western blot showed that CD38 expression was increased in peri-infarct cortex at days 1 to 7 after stroke. **b**, *Cd38* siRNA or a scrambled control was injected into lateral ventricles at 5 days after stroke. Western blot analysis confirmed that CD38 expression was successfully decreased in peri-infarct cortex at 7 days. **c**, In peri-infarct cortex, CD8 T-cell and Iba1-positive microglia/macrophages were detected by immunohistochemistry. **d**, Quantification of the number of CD8-positive or Iba1-positive cells indicated that there was no difference between control siRNA or *Cd38* siRNA ($n = 6$ mice per group). Data are mean ± s.e.m. **e**, Cultured rat cortical astrocytes were subjected to OGD for 2 h followed by treating with control siRNA or *Cd38* siRNA. Astrocyte cell morphology or GFAP expression was assessed by immunocytochemistry or western blot after 22 h reoxygenation. **f**, Morphology change was not clearly observed in cultured astrocytes treated with *Cd38* siRNA compared to control siRNA. **g**, Western blot analysis showed that CD38 was successfully decreased by siRNA transfection but GFAP expression was not changed.



Extended Data Figure 9 | Neuronal purity confirmed by FACS analysis *in vivo*. To confirm our FACS findings, we used two different published standard approaches[32,33]. **a**, By FACS, the MAP2-positive population was gated and further assessed by markers such as Iba1 (microglia/ macrophage) and GFAP (astrocyte) in brain cell samples isolated from C57BL/6 J mice. These comparisons confirmed that the $MAP2^{+}$ population did not contain any appreciable amounts of microglia or astrocytes, whereas another neuron marker (NeuN) was highly enriched. **b**, Similar findings were obtained using an alternative gating method to isolate neurons.



Extended Data Figure 10 | Involvement of integrin-mediated Src/Syk signalling mechanisms in astrocytic mitochondrial entry into neurons *in vitro*. **a**, **b**, Cultured rat cortical astrocytes were stimulated by cADPR (1 μM) for 24 h. Intracellular mitochondria labelled by MitoTracker dye was significantly increased in astrocytes stimulated with cADPR (1 μM) ($n = 2$ biological replicates, $n = 7$ independent experiments). $^{**}P < 0.01$ versus 0 h. **c**, Some mitochondria were found outside of cells. **d**, FACS analysis revealed that approximately 5×10^5 mitochondria were contained in 1 ml ACM. cADPR (1 μM) significantly increased the number of mitochondria in the media ($n = 2$ biological replicates, $n = 6$ independent experiments). **e**, Experimental schedule to quantify the entry of mitochondria into neurons after OGD. Rat cortical neurons (1×10^5 cells per well) were prepared in 24-well culture plates. ACM or cADPR-ACM (each 1 ml) was co-incubated with neurons for 18 h. Mitochondrial entry into neurons was calculated by mitochondrial intensity measured before and after washing cells with PBS. Phenol-red-free culture media was used to decrease the background signal. The background signal was subtracted from the fluorescent intensity obtained from each sample. **f**, OGD for 2 h decreased approximately 50% of the mitochondria in neurons after 18 h reoxygenation ($n = 2$ biological replicates, $n = 4$ independent experiments). **g**, Data are expressed as relative values, with total neuronal mitochondria after 2 h OGD/18 h reoxygenation being 100%. Mitochondrial entry into neurons was slightly higher after cADPR-ACM treatment (18%) compared to ACM treatment (11%), although there was no statistical significance ($n = 2$ biological replicates, $n = 4$ independent experiments). **h**, There was no difference in the percentage of mitochondrial entry after ACM treatment or cADPR-ACM treatment ($n = 2$ biological replicates, $n = 4$ independent experiments). **i**, cADPR-ACM treatment supported neuronal viability better than ACM treatment ($n = 2$ biological replicates, $n = 4$ independent experiments). **j**, Co-culture between rat cortical astrocytes in the upper chamber and rat cortical neurons in the lower chamber was performed for 18 h after OGD for 2 h in neurons. Mitochondrial entry into neurons was then measured. **k**, Immediately after OGD, dynasore (5 μM), RGDS peptide (H-Arg-Gly-Asp-Ser-OH; 50 μg ml^{-1}), or MNS (3,4-methylenedioxy-β-nitrostyrene; 1 μM) was initially added to neurons for 30 min, then astrocyte co-culture was performed for 18 h. Data are expressed as relative values, with astrocytic extracellular mitochondria plus entered mitochondria into neurons being 100%. RGDS peptide and MNS significantly decreased mitochondrial entry into neurons, but dynasore did not inhibit the entry. **l**, MNS treatment significantly decreased astrocyte-mediated neuroprotection ($n = 2$ biological replicates, $n = 4$ independent experiments). **m**, Dynasore (5 μM), RGDS peptide (50 μg ml^{-1}), or MNS (1 μM) did not affect neuronal viability after 2 h OGD ($n = 4$). Data are mean ± s.e.m. These data suggest that the entry of astrocytic mitochondrial particles into neurons may involve integrin-mediated Src/Syk signalling mechanisms. However, we acknowledge that these pathways may be multifactorial, and deeper analyses are warranted to dissect entry mechanisms under various physiological and pathological conditions. *P* values are from a one-way ANOVA followed by Tukey's test (**a, i, k, l**) or an unpaired *t*-test (**d, g**).

LETTER

doi:10.1038/nature18929

HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption

Johannes F. Scheid[1,2]*, Joshua A. Horwitz[1]*, Yotam Bar-On[1], Edward F. Kreider[3], Ching-Lan Lu[1], Julio C. C. Lorenzi[1], Anna Feldmann[4], Malte Braunschweig[1], Lilian Nogueira[1], Thiago Oliveira[1], Irina Shimeliovich[1], Roshni Patel[1], Leah Burke[5], Yehuda Z. Cohen[1], Sonya Hadrigan[1], Allison Settler[1], Maggi Witmer-Pack[1], Anthony P. West Jr[6], Boris Juelg[7], Tibor Keler[8], Thomas Hawthorne[8], Barry Zingman[9], Roy M. Gulick[5], Nico Pfeifer[4], Gerald H. Learn[3], Michael S. Seaman[10], Pamela J. Bjorkman[6], Florian Klein[1,11,12], Sarah J. Schlesinger[1], Bruce D. Walker[7,13], Beatrice H. Hahn[3], Michel C. Nussenzweig[1,14] & Marina Caskey[1]

Interruption of combination antiretroviral therapy in HIV-1-infected individuals leads to rapid viral rebound. Here we report the results of a phase IIa open label clinical trial evaluating 3BNC117, a broad and potent neutralizing antibody against the CD4 binding site of the HIV-1 Env protein[1], during analytical treatment interruption in 13 HIV-1-infected individuals. Participants with 3BNC117-sensitive virus outgrowth cultures were enrolled. Results show that two or four 30 mg kg^{-1} 3BNC117 infusions, separated by 3 or 2 weeks, respectively, are generally well tolerated. Infusions are associated with a delay in viral rebound of 5–9 weeks after two infusions, and up to 19 weeks after four infusions, or an average of 6.7 and 9.9 weeks, respectively, compared with 2.6 weeks for historical controls ($P < 0.00001$). Rebound viruses arise predominantly from a single provirus. In most individuals, emerging viruses show increased resistance, indicating escape. However, 30% of participants remained suppressed until antibody concentrations waned below 20 μg ml^{-1}, and the viruses emerging in all but one of these individuals showed no apparent resistance to 3BCN117, suggesting failure to escape over a period of 9–19 weeks. We conclude that the administration of 3BNC117 exerts strong selective pressure on HIV-1 emerging from latent reservoirs during analytical treatment interruption in humans.

A fraction of HIV-1-infected individuals develops broad and potent serologic activity against the virus. Single-cell antibody cloning methods[2] have uncovered the source of this activity as broadly neutralizing antibodies (bNAbs), which target different sites on the HIV-1 envelope spike protein, gp160 (refs 1–3).

In animal models, bNAbs show potent prophylactic activity, suppress established viraemia, and delay viral rebound during analytical treatment interruption (ATI)[4–8]. In humans, a phase I clinical trial showed that 3BNC117 is generally safe and effective in transiently reducing viraemia in chronically HIV-1-infected individuals[9]. A single infusion of 3BNC117 was well tolerated, rapidly decreased viral loads in viraemic individuals by an average of 1.48 $\log_{10}$ copies per ml, with durable activity for 4 weeks[9]. In addition, 3BNC117 increased autologous antibody responses in HIV-1-infected individuals, and enhanced clearance of infected cells in humans and in humanized mice[10,11]. VRC01, a less potent bNAb that also targets the CD4-binding site, suppressed viraemia by 1.14 $\log_{10}$ (refs 12, 13 and Fig. 1a, b).

To investigate whether 3BNC117 can suppress viral rebound from the latent reservoir during ATI in chronically suppressed HIV-1 infected humans, we conducted a phase IIa open label clinical trial. To select participants with 3BNC117-sensitive viruses in their latent reservoirs, we performed bulk viral outgrowth cultures of peripheral blood mononuclear cells (PBMCs) from individuals whose viraemia was suppressed by combination antiretroviral therapy (ART). The resulting isolates were screened for sensitivity to 3BNC117 using the TZM-bl assay (Supplementary Table 1). Of 63 individuals screened, only 11% yielded viruses that were fully resistant to 3BNC117 ($IC_{50} > 20$ μg/ml), and 65% were sensitive to 3BNC117 IC_{50} at concentrations below 2.0 μg/ml. In contrast only 29% were similarly sensitive to VRC01 (Fig. 1a and b, Extended Data Fig. 1 and Supplementary Table 1).

We enrolled HIV-1 infected individuals who were on suppressive antiretroviral therapy (ART) with plasma viral loads <50 HIV-1 RNA copies per ml for at least 12 months, had CD4 counts >500 cells per mm^3, yielded 3BNC117-sensitive outgrowth viruses ($IC_{50} \leq 2.0$ μg ml^{-1}), and whose viral load at screen was <20 copies per ml (Extended Data Fig. 1, Supplementary Tables 2–4, and Methods). Participants were enrolled in two groups: eight in group A to receive two 30 mg kg^{-1} infusions three weeks apart, while seven in group B received up to four 30 mg kg^{-1} infusions at two-week intervals (Fig. 1c, d, Supplementary Table 2). Two group A participants had viral loads >20 copies per ml at the time of infusion and were excluded from further analysis (Supplementary Tables 2 and 4). Participants are numbered 701–715 (Supplementary Table 2).

ATI was started 2 days after the first 3BNC117 infusion. ART was reinitiated and infusions were stopped after two consecutive plasma viral load measurements exceeded 200 copies per ml. All individuals on non-nucleoside reverse transcriptase inhibitors (NNRTIs) were switched to an integrase-inhibitor-based regimen (dolutegravir plus tenoforvir disoproxil fumarate/emtricitabine) four weeks before ATI owing to the long half-life of NNRTIs (Supplementary Table 2).

Both dosing regimens were generally well tolerated. The majority of reported adverse events were transient and grade 1 in severity (Supplementary Table 5). The mean CD4 T-cell count at baseline (day 0) was 747 cells per mm^3, and the average change in CD4 T-cell counts between start of ATI and rebound was −127 cells per mm^3. Although CD4 T cells declined modestly during viral rebound in some participants, CD4 T-cells returned to baseline by week 12 in most participants (mean 828 cells per mm^3) (Extended Data Fig. 2 and Supplementary Table 4). Of 12 individuals tested, 5 showed measurable increases in the magnitude and/or breadth of T cell responses to HIV-1 12 weeks

[1]Laboratory of Molecular Immunology, The Rockefeller University, New York, New York 10065, USA. [2]Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. [3]Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [4]Department of Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, Campus E1 4, 66123 Saarbrücken, Germany. [5]Division of Infectious Diseases, Weill Medical College of Cornell University, New York, New York 10065, USA. [6]Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. [7]Ragon Institute of MGH, MIT and Harvard, Cambridge, Massachusetts 02139, USA. [8]Celldex Therapeutics, Inc., Hampton, New Jersey 08827, USA. [9]Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, New York 10467, USA. [10]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. [11]Laboratory of Experimental Immunology, Center for Molecular Medicine Cologne (CMMC), University of Cologne, 50931 Cologne, Germany. [12]Department I of Internal Medicine, Center of Integrated Oncology Cologne-Bonn, University Hospital Cologne, 50937 Cologne, Germany. [13]Howard Hughes Medical Institute, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [14]Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA.
*These authors contributed equally to this work.

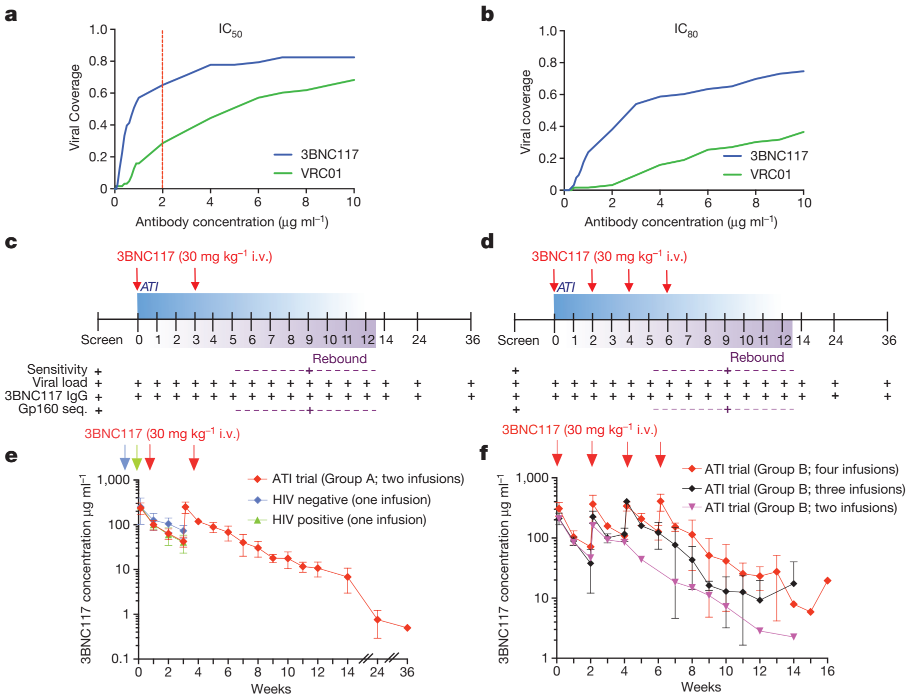


Figure 1 | 3BNC117 neutralization coverage, trial design and pharmacokinetics of 3BNC117 in HIV-1-infected individuals during ATI. **a**, **b**, Sensitivity of virus outgrowth cultures from 63 ART suppressed individuals to 3BNC117 and VRC01 (Supplementary Table 1). The *y*-axis shows the fraction of viral outgrowth culture supernatants neutralized by a given antibody concentration (*x*-axis) in Tzm-bl assays. Red line indicates cut-off IC_{50} (2 μg ml^{-1}) for participation in the trial. **c**, **d**, Diagrammatic representation of study groups A and B respectively. 3BNC117 infusions indicated by the red arrows, and sampling for PK and virologic studies indicated below. Numbers indicate study weeks. **e**, **f**, 3BNC117 levels as determined by ELISA for group A ($n = 6$, left panel, red), group B ($n = 7$, right panel, red ($n = 4$), black ($n = 2$) and purple ($n = 1$)), HIV-1 negative ($n = 3$, blue) and viraemic individuals ($n = 6$, green)[9]. Curves indicate mean 3BNC117 levels, error bars indicate standard deviation. Arrows indicate 3BNC117 infusions.

after ATI, relative to baseline (Extended Data Fig. 3). None of the participants experienced acute retroviral syndrome during rebound, and viraemia was re-suppressed below 20 copies per ml in all participants within 2–7 weeks after restarting ART (Supplementary Table 4). We conclude that up to four 30 mg kg^{-1} infusions of 3BNC117 during ATI are generally safe and well tolerated.

By anti-idiotype ELISA[9] the half-life of 3BNC117 during ATI was 19.6 days among group A participants, and 14.1 days among those in group B (Fig. 1e, f and Supplementary Table 4). These measurements are similar to previously reported values for 3BNC117 in HIV-1-infected individuals on ART[9] (Fig. 1e).

All six group A participants maintained viral loads below 200 copies per ml during the first 4 weeks, with rebound 5–9 weeks after ART interruption (Fig. 2a and Supplementary Table 4a). In group B, rebound occurred 3–19 weeks after ATI, with four out of seven (57%) participants remaining suppressed for at least 10 weeks (Fig. 2b and Supplementary Table 4b). The average time to rebound was 6.7 weeks in group A, 9.9 weeks in group B, and 8.4 weeks for all participants together, compared with 2.6 weeks for matched historical non-infused control individuals (Fig. 2c, Extended Data Fig. 4a, Supplementary Tables 4, 6 and 7). Altogether, 6 of the 13 infused individuals (46%) remained suppressed until at least 9 weeks after ATI. Relative to matched historical control individuals, the delay to rebound among all 3BNC117-infused participants was highly significant ($P < 0.00001$ weighted log-rank test, Fig. 2c, Extended Data Fig. 4, Supplementary Tables 4, 6 and 7 and Methods). We conclude that repeated infusions of 3BNC117 are generally safe and significantly delay HIV-1 rebound from the latent reservoir during ATI.

Time to viral rebound did not correlate with pre-ATI viral culture sensitivity to 3BNC117, nor to baseline levels of cell-associated HIV-1 DNA (Fig. 2d and Extended Data Fig. 4e). Therefore the significance of viral outgrowth sensitivity as an inclusion criterion is not clear. 3BNC117 levels at rebound were also variable, ranging from 6–168 μg ml^{-1}, but directly correlated with the IC_{80} of the emerging virus (Fig. 2a, b, e).

To determine whether rebound was associated with resistance to 3BNC117, we compared pre-infusion and rebound viral outgrowth cultures. A majority (8/13) of participants had rebound viruses that were more resistant to 3BNC117 (IC_{80} > threefold higher, Fig. 3a, c, Extended Data Fig. 5a, Supplementary Table 3). Among group A participants, all but one (707) had more resistant rebound viruses; however, among group B participants, four of seven (710, 711, 712 and 715) showed similar pre-infusion and rebound sensitivity to 3BNC117 (Fig. 3a, c, Extended Data Fig. 5a, Supplementary Table 3). Among these five individuals, 711 was the earliest to rebound at 3 weeks, despite having viruses that were surprisingly sensitive to 3BNC117 as measured by IC_{50} (Fig. 2b, Extended Data Fig. 5a, Supplementary Tables 3 and 4). However, 100% neutralization was not achieved against 711 rebound or pre-infusion viruses, even at high (50 μg ml^{-1}) antibody concentrations, suggesting that 3BNC117 was not fully therapeutic (Extended Data Fig. 5a, Supplementary Table 3). Thus, the only participant in the study to rebound within 3 weeks of ATI may have done so because of pre-existing resistance to 3BNC117 by the dominant virus in the reservoir.

The other four participants that showed no change between pre- and post-infusion culture sensitivity to 3BNC117, 707, 710, 712, and 715 rebounded relatively late at 9, 19, 16 and 11 weeks after ATI, respectively (Figs 2a, b, 3a, c, Extended Data Fig. 5, Supplementary Tables 3 and 4). In all of these individuals rebound was associated with relatively low antibody concentrations ranging from 6–41 μg ml^{-1} (mean 19.7 μg ml^{-1}). This antibody concentration represents 9.6-fold the mean IC_{80} for the rebounding viruses, which is consistent with previous reports on the relationship between suppressive 3BNC117 concentration and neutralization titre in macaques[14] (Fig. 2, Extended Data Fig. 5, Supplementary Tables 3 and 4).

To determine whether viral rebound during ATI was associated with resistance to other bNAbs undergoing clinical testing, we examined sensitivity to 10-1074 (ref. 15), which targets a different and non-overlapping epitope on the HIV-1 trimer (Fig. 3b, d, Extended Data Fig. 5, and Supplementary Table 3). With the exception of 703 and 711, the participants' rebound cultures did not show increased resistance to 10-1074. We conclude that rebound during ATI in the presence of 3BNC117 is infrequently associated with increased resistance to 10-1074.

To characterize viruses emerging from the latent reservoir further, we performed single genome sequencing (SGS) of viral RNA from the plasma and viral outgrowth cultures from eight individuals. Phylogenetic analysis of these sequences indicated that all of these eight trial participants were infected with epidemiologically unrelated viruses (Extended Data Fig. 6). Given the limited sampling of the pre-infusion reservoir, rebound viruses did not always fall within the radiation

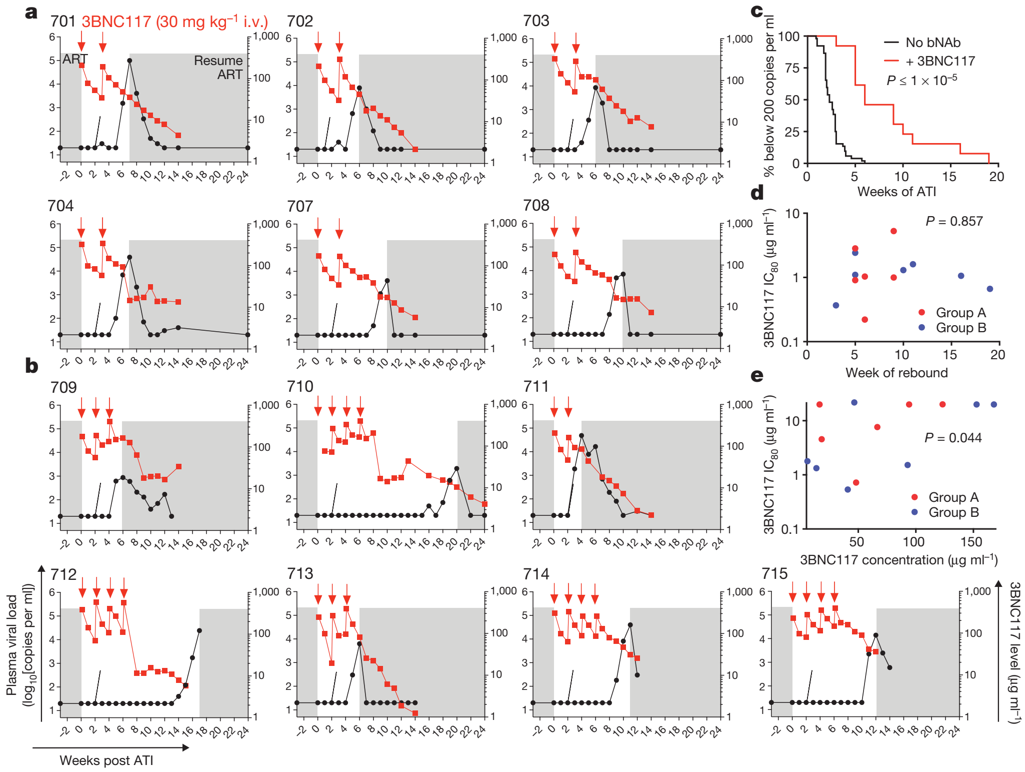


Figure 2 | Delay in viral rebound in the presence of 3BNC117. **a**, **b**, Plasma viral loads and 3BNC117 levels in group A and group B participants respectively. 3BNC117 infusions are indicated with red arrows. The left *y*-axis shows plasma viral loads in RNA copies per ml (black curves), and right *y*-axis shows antibody levels measured by ELISA (red curves). Average rebound time point (2.6 weeks, Supplementary Table 6) in 52 ACTG trial participants who underwent ATI without antibody treatment[29] is shown with dotted lines. Grey areas indicate ART therapy. **c**, Kaplan–Meier plot summarizing viral rebound in 52 ACTG trial participants who underwent ATI without antibody treatment (black, Supplementary Table 6), and the combination of all 13 participants (red) who underwent ATI with 3BNC117 infusions. The *y*-axis indicates the percentage of participants with viral levels below 200 RNA copies per ml, *x*-axis indicates weeks after ATI initiation. The *P* value is based on a bootstrap version of the weighted log-rank test adjusting for the potential confounders 'years on ART', and 'age' (Supplementary Table 7, Methods Statistical Analyses). **d**, Dot plot indicating the relationship between 3BNC117 sensitivity of pre-infusion outgrowth cultures at screening (*y*-axis, IC$_{80}$ in μg ml^{-1}) and the week of rebound (*x*-axis). Group A ($n = 6$) and group B ($n = 7$) participants are coloured red and blue respectively. The *P* value was derived from calculating the Pearson correlation coefficients. **e**, Dot plot indicating the relationship between 3BNC117 sensitivity of rebound outgrowth cultures (*y*-axis, IC$_{80}$ in μg ml^{-1}) and the 3BNC117 serum concentration at rebound (*x*-axis, in μg ml^{-1}). 704, 708, 709 and 713 did not reach IC$_{80}$ at the concentrations tested and were assigned a value of 22 μg ml^{-1}. Group A ($n = 6$) and group B ($n = 7$) participants are coloured red and blue, respectively. The *P* value was derived from calculating the Pearson correlation coefficients.

of pre-infusion viral isolates (Fig. 3e, f, Extended Data Figs 7 and 8, Supplementary Figs 1 and 2).

Remarkably, in five of eight participants, all rebounding virus sequences clustered within a low diversity lineage, consistent with the clonal expansion of a single recrudescent virus (Fig. 3e, f, Extended Data Figs 7 and 8, Supplementary Table 8). These data contrast with individuals undergoing ATI in the absence of antibody infusion, where virus rebound is consistently polyclonal, indicating the activation of multiple latently infected cells[16–19]. Thus, in addition to delaying rebound, 3BNC117 appears to restrict the outgrowth of viral genotypes from the latent reservoir.

Six of eight participants sequenced had rebound viral outgrowth culture and/or plasma sequences that indicated 3BNC117 resistance. For example, in 704, all rebound viruses carried a serine at position 456 (Fig. 4a and Supplementary Figs 1 and 2), which may disrupt a highly conserved salt bridge that maintains the V5 loop's position and conformation[20–22]. Similarly, in 708 and 709, nearly all rebound viruses carried atypical residues at position 282, where a lysine residue typically forms a salt bridge with 3BNC117 (Fig. 4a and Supplementary Figs 1 and 2)[23]. However, documented 3BNC117 resistance mutations[24] were not universally identified among rebound viral strains (Fig. 4a and Supplementary Figs 1 and 2). Only a minor fraction (3 of 23) of sequences in the rebound population of participant 701 had potential resistance-conferring residues in Loop D (274F, 282R), while the remaining sequences did not (Fig. 4a and Supplementary Figs 1 and 2). Similarly, in 702 and 703, only a subset of rebound viruses carried a putative resistance-conferring A281D change[1,23]. Nevertheless, the frequency of this change increased markedly with time in both participants, indicating continued selection for 3BNC117 resistance (Fig. 4a and Supplementary Figs 1 and 2). For participants 707 and 711, no sequence features were identified that would indicate 3BNC117 resistance.

To determine the sensitivity of rebound viruses to 3BNC117, we performed TZM-bl neutralization assays using pseudoviruses typed with SGS Env genotypes (Fig. 4b, Extended Data Figs 7 and 8, Supplementary Figs 1 and 2 and Supplementary Table 9). With the exception of participant 707, who rebounded 9 weeks after ATI at very low 3BNC117 titres, Env genotypes at rebound were more resistant to 3BNC117 than pre-ATI (Fig. 4b, Extended Data Figs 7 and 8 and Supplementary Table 9). We conclude that viral rebound during ATI in the presence of 3BNC117 selects for the emergence of resistant variants, indicating strong selection pressure by this antibody on viral populations arising from the reservoir.

Antibody potency and half-life are directly correlated with HIV-1 prophylaxis in pre-clinical models. For example, VRC01, a CD4bs antibody that is less potent than 3BNC117[1], is less effective than 3BNC117 in preventing SHIV$_{AD8}$ infection in macaques[8,25]. Consistent with these observations, clinical trials with combinations of three less-potent

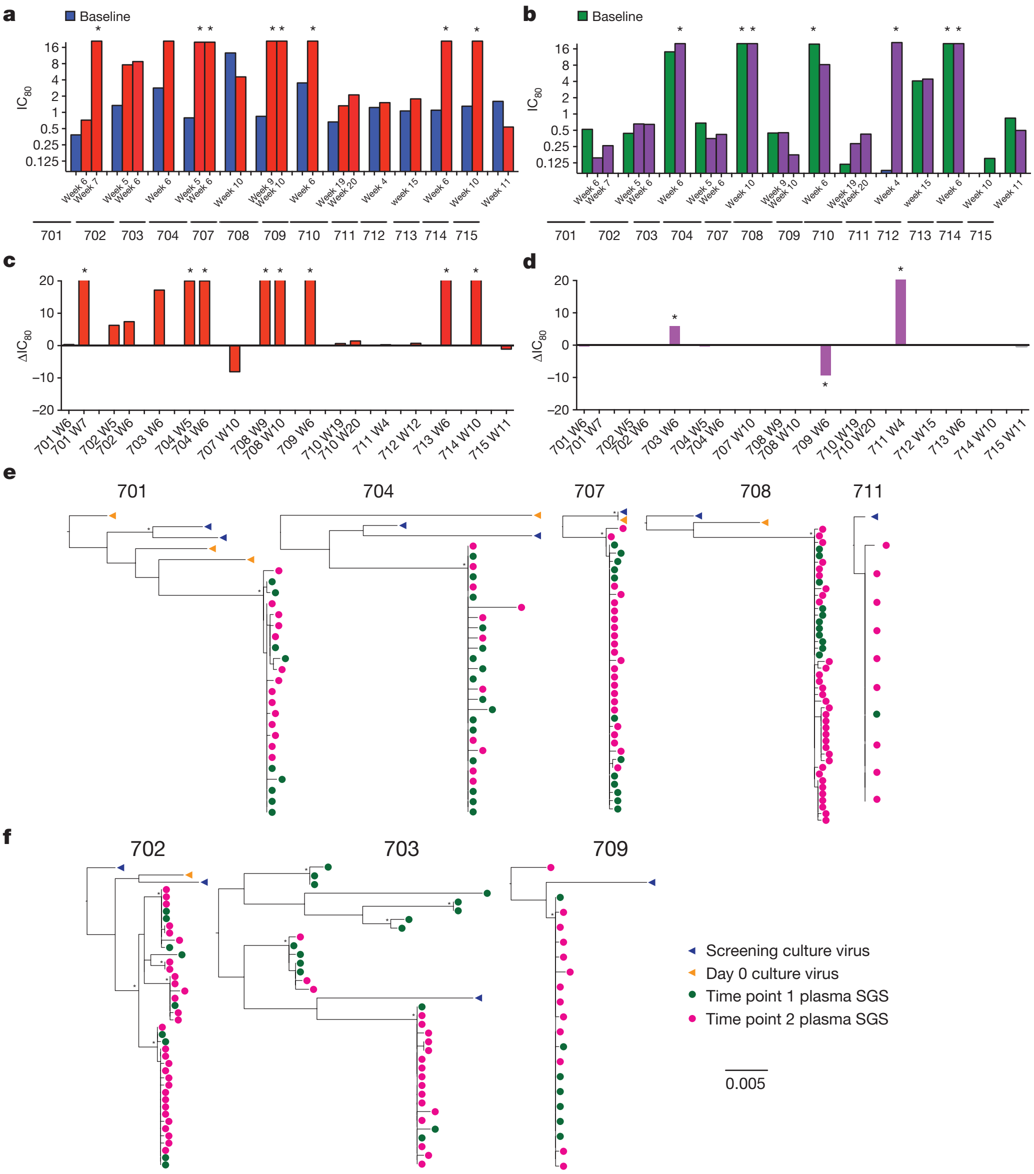


Figure 3 | Viral rebound during ATI and 3BNC117 treatment. **a**, **b**, Graph of 3BNC117 (**a**) or 10-1074 (**b**) IC_{80} titres against baseline and rebound outgrowth cultures. Blue and green bars represent average IC_{80} titres against screen and day 0 outgrowth cultures; red and purple bars represent IC_{80} titres against rebound outgrowth cultures from the indicated time points. Asterisks indicate cultures failing to reach an IC_{80} up to $20\,\mu g\,ml^{-1}$. **c**, **d**, Difference between rebound and pre-infusion culture IC_{80} titres (from **a**, **b**) for 3BNC117 (**c**) or 10-1074 (**d**). For cultures failing to reach an IC_{80} up to $20\,\mu g\,ml^{-1}$, a value of $20\,\mu g\,ml^{-1}$ was assigned, and such cultures are marked with an asterisk. **e**, **f**, Clonality of the rebound virus. Maximum likelihood phylogenetic trees comparing pre-ATI single genome derived *env* sequences (blue and orange) to rebound plasma *env* sequences (green and pink) are shown for participants whose rebound comprised single (**e**) versus multiple (**f**) viruses (Supplementary Table 8). Pre-ATI culture sequences were inferred as described in the Methods section 'Statistical analyses'.

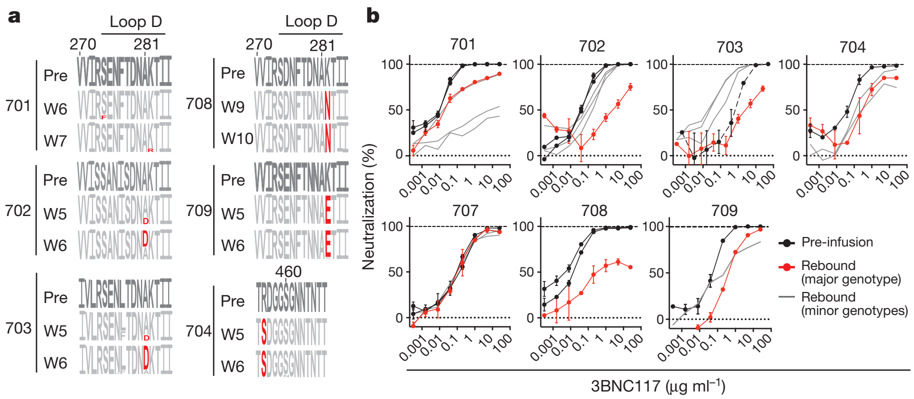


Figure 4 | 3BNC117 resistance in rebound viruses. **a**, Logogram shows *env* gp120 regions (amino acid positions; 270–285 and 455–467, according to HXBc2 numbering) indicating sequence changes from pre-infusion culture(s) (first row) to rebound sequences derived from plasma SGS at the indicated time points. The frequency of each amino acid is indicated by its height. Red residues represent mutations predicted to affect neutralization[30]. **b**, 3BNC117 neutralization sensitivity of pseudoviruses derived from pre-infusion or rebound SGS. Black lines represent pre-infusion virus *envs*; red lines represent the major *env* at rebound for each participant (Extended Data Figs 7 and 8, Supplementary Tables 8 and 9 and Methods); grey lines represent minor rebound *envs* in participants with multiple rebound viruses or variants that evolved after rebound (Extended Data Figs 7 and 8, Supplementary Table 9). Symbols reflect the means of two technical replicates; error bars denote standard deviation.

first-generation bNAbs showed limited effects on viral rebound in the setting of ATI in chronically infected individuals[26,27]. In addition, selective pressure as evidenced by escape mutations was only observed for one of the three antibodies used in the combination, 2G12[26,27]. In contrast, 3BNC117 alone significantly delayed viral rebound with nearly half of all individuals remaining below 200 copies per ml until at least 9 weeks, including four individuals who failed to develop resistance and only rebounded at low antibody concentrations. We speculate that the difference in efficacy between 3BNC117 and less potent bNAbs in the setting of ATI is due to increased potency and/or a longer half-life[1,9,13].

Nevertheless, the majority of the individuals we studied rebounded at high 3BNC117 serum concentrations. A single viral genotype displaying increased resistance to 3BNC117 established rebound in most cases. These viruses represent pre-existing dormant variants that emerged from the latent reservoir. The time to rebound did not correlate with the amount of viral DNA in circulating PBMCs; however, this is a poor measure of the HIV-1 reservoir, since most integrated proviruses in patients on ART are defective[28]. Instead, the delay in viral rebound may represent a measure of the frequency of 3BNC117-resistant variants in the latent reservoir.

Combinations of drugs are needed to maintain HIV-1 suppression in effective ART regimens. Similarly, combinations of antibodies were required to suppress viraemia in humanized mice[6,7]. We speculate that combinations of bNAbs will also be needed to increase the frequency of individuals that remain suppressed by antibody during ATI.

Whether 3BNC117 can also impact the size and composition of the latent reservoir during ATI will require additional studies.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 May; accepted 15 June 2016.
Published online 22 June 2016.

1. Scheid, J. F. *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* **333,** 1633–1637 (2011).
2. Scheid, J. F. *et al.* Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458,** 636–640 (2009).
3. Klein, F. *et al.* Antibodies in HIV-1 vaccine development and therapy. *Science* **341,** 1199–1204 (2013).
4. Barouch, D. H. *et al.* Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. *Nature* **503,** 224–228 (2013).
5. Halper-Stromberg, A. *et al.* Broadly neutralizing antibodies and viral inducers decrease rebound from HIV-1 latent reservoirs in humanized mice. *Cell* **158,** 989–999 (2014).
6. Klein, F. *et al.* HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature* **492,** 118–122 (2012).
7. Horwitz, J. A. *et al.* HIV-1 suppression and durable control by combining single broadly neutralizing antibodies and antiretroviral drugs in humanized mice. *Proc. Natl Acad. Sci. USA* **110,** 16538–16543 (2013).
8. Shingai, M. *et al.* Passive transfer of modest titers of potent and broadly neutralizing anti-HIV monoclonal antibodies block SHIV infection in macaques. *J. Exp. Med.* **211,** 2061–2074 (2014).
9. Caskey, M. *et al.* Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* **522,** 487–491 (2015).
10. Schoofs, T. *et al.* HIV-1 therapy with monoclonal antibody 3BNC117 elicits host immune responses against HIV-1. *Science* **352,** 997–1001 (2016).
11. Lu, C. L. *et al.* Enhanced clearance of HIV-1-infected cells by broadly neutralizing antibodies against HIV-1 *in vivo*. *Science* **352,** 1001–1004 (2016).
12. Kong, R. *et al.* Improving neutralization potency and breadth by combining broadly reactive HIV-1 antibodies targeting major neutralization epitopes. *J. Virol.* **89,** 2659–2671 (2015).
13. Lynch, R. M. *et al.* Virologic effects of broadly neutralizing antibody VRC01 administration during chronic HIV-1 infection. *Sci. Transl. Med.* **7,** 319ra206 (2015).
14. Shingai, M. *et al.* Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia. *Nature* **503,** 277–280 (2013).
15. Mouquet, H. *et al.* Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl Acad. Sci. USA* **109,** E3268–E3277 (2012).
16. Rothenberger, M. K. *et al.* Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. *Proc. Natl Acad. Sci. USA* **112,** E1126–E1134 (2015).
17. Kearney, M. F. *et al.* Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. *PLoS Pathog.* **10,** e1004010 (2014).
18. Kearney, M. F. *et al.* Origin of rebound plasma HIV includes cells with identical proviruses that are transcriptionally active before stopping of antiretroviral therapy. *J. Virol.* **90,** 1369–1376 (2015).
19. Salantes, B. S. B; Bar, Katharine. Potency and Kinetics of Autologous HIV-1 Neutralizing Antibody Responses During ATI. *CROI Conference Abstracts* **Abstract #92** (2016).
20. West, A. P., Jr, Diskin, R., Nussenzweig, M. C. & Bjorkman, P. J. Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc. Natl Acad. Sci. USA* **109,** E2083–E2090 (2012).
21. Diskin, R. *et al.* Restricting HIV-1 pathways for escape using rationally designed anti-HIV-1 antibodies. *J. Exp. Med.* **210,** 1235–1249 (2013).
22. Lynch, R. M. *et al.* HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. *J. Virol.* **89,** 4201–4213 (2015).
23. Zhou, T. *et al.* Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. *Immunity* **39,** 245–258 (2013).
24. Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342,** 1484–1490 (2013).
25. Gautam, R. *et al.* A single injection of anti-HIV-1 antibodies protects against repeated SHIV challenges. *Nature* **533,** 105–109 (2016).
26. Trkola, A. *et al.* Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. *Nat. Med.* **11,** 615–622 (2005).
27. Mehandru, S. *et al.* Adjunctive passive immunotherapy in human immunodeficiency virus type 1-infected individuals treated with antiviral therapy during acute and early infection. *J. Virol.* **81,** 11016–11031 (2007).
28. Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155,** 540–551 (2013).
29. Li, J. Z. *et al.* The size of the expressed HIV reservoir predicts timing of viral rebound after treatment interruption. *AIDS* **30,** 343–353 (2016).
30. West, A. P., Jr *et al.* Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc. Natl Acad. Sci. USA* **110,** 10598–10603 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank the trial participants for their invaluable support. We thank the Rockefeller University Hospital Clinical Research Support Office and nursing staff for help with recruitment and study implementation, especially N. Buckley, A. Hurley, S. B. A. Shulman and L. Corregano; all members of the Nussenzweig laboratory, especially T. Schoofs, A. Halper-Stromberg, M. and Z. Jankovic; C. Unson-O'Brien, J. Dizon, R. Baptiste and R. Levin for sample processing and study coordination; A. Louie for regulatory support; P. Fast and H. Park for clinical monitoring; E. Giorgi and W. Fischer from Los Alamos National Laboratory; R. T. Gandhi, J. Li and The AIDS Clinical Trials Group (grant UM1 AI068636) and its Statistical and Data Management Center (grant UM1 AI068634). This study was supported by the following grants: Collaboration for AIDS Vaccine Discovery grant OPP1033115 (M.C.N.) and OPP1032144 (M.S.S.). Grant 8 UL1 TR000043 from the National Center for Advancing Translational Sciences (NCATS); NIH Clinical and Translational Science Award (CTSA) program; NIH Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID) 1UM1 AI100663-01 (M.C.N.) and 5UM1 AI100645-03 (B.H.H.); Bill and Melinda Gates Foundation grants OPP1092074 and OPP1124068 (M.C.N.); NIH HIVRAD P01 AI100148 (P.J.B. and M.C.N.); the Robertson Foundation to M.C.N. M.C.N. is a Howard Hughes Medical Institute Investigator. Ruth L. Kirschstein National Research Service Award F30 AI112426 (E.F.K.); F31 AI118555 (J.A.H.); The NIH Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID) 1UM1 AI00645 (B.H.H.); The University of Pennsylvania Center for AIDS Research (CFAR) Single Genome Amplification Service Center P30 AI045008 (B.H.H.); The NIH Scripps Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID and 1UM1-AI100663) (B.D.W.).

Author Contributions M.C.N., J.F.S., J.A.H. and M.C. wrote the manuscript; J.F.S., M.C. and M.C.N. designed the trial; J.F.S., J.A.H., Y.B., J.C.C.L., L.N., Y.Z.C., C.-L.L. and M.B. performed tissue culture experiments and SGS amplifications; M.S.S. performed TZM-bl assays; J.F.S., J.A.H., Y.B., E.F.K., T.O., A.P.W., G.H.L., P.J.B., F.K., S.J.S., B.H.H., M.C.N. and M.C. analysed the data; E.F.K., G.H.L. and B.H.H. performed SGA analysis; I.S., R.P. and J.F.S. processed patient samples; L.B., S.H., A.S., M.W.-P., B.Z., R.M.G., S.J.S. and M.C. performed patient recruitment; A.F. and N.P. performed statistical analyses; B.J. and B.D.W. performed antigen-specific T cell experiments; T.K. and T.H. produced 3BNC117 and provided PK data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C.N. (nussen@rockefeller.edu) or M.C. (mcaskey@rockefeller.edu).

Reviewer Information *Nature* thanks S. Deeks, D. Richman and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Study design. An open-label, dose-escalation phase 2a study was conducted in HIV-1-infected participants (http://www.clinicaltrials.gov; NCT02446847). Study participants were enrolled sequentially according to eligibility criteria. Group A received 3BNC117 on days 0 and 21 at a dose of 30 mg/kg body weight at a rate of 250 ml/hour. Group B received 3BNC117 on days 0, 14, 28 and 42 at a dose of 30 mg/kg, as long as viral rebound did not occur. Antiretroviral therapy (ART) was discontinued 2 days after the first 3BNC117 infusion (day 2). Plasma HIV-1 RNA levels were monitored weekly, and ART was resumed when viral load increased to ≥200 c.p.m. in two consecutive weekly measurements.

Study participants were followed for 36 weeks after the first infusion. Safety data are reported until week 36 for participants enrolled in group A and until week 14 for participants enrolled in group B. All participants provided written informed consent before participation in the study and the study was conducted in accordance with Good Clinical Practice. The protocol was approved by the Federal Drug Administration in the USA and the Institutional Review Board at the Rockefeller University.

Study participants. All study participants were recruited at the Rockefeller University Hospital, New York, USA. Eligible participants were adults aged 18–65 years, HIV-1-infected and before enrolment had plasma HIV-1 RNA levels <50 c.p.m. for at least 12 months while on combination ART and <20 c.p.m. at the screening visit, and current CD4 count >500/μl. In addition, participant-derived HIV-1 isolates produced by co-culture of participant PBMCs with HIV-uninfected donor PBMCs were required to be neutralized by 3BNC117 with an IC_{50} <2 μg/ml in TZM-bl neutralization assays, as previously described[31]. An IC_{50} of <2 ug/ml was chosen as a cut-off based on previous PK data of 3BNC117 in humans[9] and data in macaques showing that antibody levels 10–100 times the IC_{50} value against infecting viral strains are necessary to control viral rebound[14]. However, given the limited diversity and representation of the latent reservoir in outgrowth cultures (Supplementary Fig. 2) and the fact that no correlation between pre-infusion IC_{50} and delay of viral rebound was found in this study, the significance of this criterion is unclear. Participants on an NNRTI-based ART regimen were switched to a study-provided integrase-inhibitor-based regimen (dolutegravir (Tivicay, ViiV Pharmaceuticals) + tenofovir disoproxil fumarate/emtricitabine (Truvada, Gilead Sciences) 4 weeks before treatment interruption due to the prolonged half-life of NNRTIs. Exclusion criteria included history of CD4 nadir <200 cells/μl, concomitant hepatitis B or C infections, previous receipt of a monoclonal antibody of any kind, or clinically relevant physical findings, medical conditions or laboratory abnormalities. Pregnant and breastfeeding women were not eligible.

Historical controls (ACTG trial participants). Viral rebound data from 52 participants who participated in four ACTG ATI studies without additional interventions (ACTG 371[32], A5024[33], A5068[34], and A5197[32]) were compared with viral rebound data in this study. Historical controls were selected based on similar inclusion criteria: age 18–65, Plasma HIV-1 RNA <50 c.p.m. for at least 12 months before ATI while on combination ART, CD4 count at time of ATI >500 cells/μl, CD4 nadir >200 cells/μl, weekly viral load measurements at least until viral rebound occurred.

Study procedures. The appropriate volume of 3BNC117 was calculated according to body weight, diluted in sterile normal saline to a total volume of 250 ml, and administered intravenously over 60 min. Study participants received 3BNC117 on days 0 and 21, or 0, 14, 28, and 42 and remained under monitoring at the Rockefeller University Hospital for 4 h after each infusion. Participants returned for frequent follow up visits for safety assessments, which included physical examination, measurement of clinical laboratory parameters such as haematology, chemistries, urinalysis, and pregnancy tests (for women). Plasma HIV-1 RNA levels were monitored weekly during the ATI period and CD4 counts were measured every other week (Supplementary Table 4). Study investigators evaluated and graded adverse events according to the DAIDS AE Grading Table and determined causality. Blood samples were collected before and at multiple times after 3BNC117 infusions. Samples were processed within 4 h of collection, and serum and plasma samples were stored at −80 °C. PBMCs were isolated by density gradient centrifugation. The absolute number of peripheral blood mononuclear cells was determined by an automated cell counter (Vi-Cell XR; Beckman Coulter), and cells were cryopreserved in fetal bovine serum plus 10% DMSO.

ART re-initiation criteria. Antiretroviral therapy was discontinued 2 days after the first 3BNC117 infusion (day 2). ART was re-initiated when HIV-1 RNA levels were found to be ≥200 c.p.m. and/or CD4 T cell counts decreased to <350 cells/μl and the result was confirmed with a repeat measurement.

Plasma HIV-1 RNA Levels. HIV-1 RNA levels in plasma were measured at the time of screening (within 49 days before the first infusion), day 0 (before infusion), and weekly until week 12, then at weeks 14, 24 and 36. Participants that remained virologically suppressed to <20 c.p.m. off ART beyond week 12, returned for weekly measurements of plasma HIV-1 RNA levels. HIV-1 RNA levels were determined using the Roche COBAS AmpliPrep/COBAS TaqMan HIV-1 Assay, Version 2.0, which detects between 20 and 1×10^7 c.p.m. This assay was performed at LabCorp.

$CD4^+$ and $CD8^+$ T cells. $CD4^+$ and $CD8^+$ T-cell counts were determined at screening, on day 0 (before infusion), and weeks 2, 3, 4, 6, 8, 10, 12, 14, and 36 by a clinical flow cytometry assay, performed at LabCorp. Cells were analysed by flow cytometry. Leukocytes were determined as $CD45^+$ cells. Percentage of cells positively stained for CD3, CD4, CD8 as well as the CD4/CD8 ratio were analysed with the BD Multiset software (BD Biosciences).

3BNC117 study drug. 3BNC117 is a recombinant, fully human IgG1κ mAb recognizing the CD4 binding site on the HIV-1 envelope[1]. The antibody was cloned from an HIV-1-infected viraemic controller in the International HIV Controller Study[1,35], expressed in Chinese hamster ovary cells (clone 5D5-5C10), and purified using standard methods. The 3BNC117 drug substance was produced at Celldex Therapeutics Fall River (MA) GMP facility, and the drug product was fill-finished at Gallus BioPharmaceuticals (NJ). The resulting purified 3BNC117 was supplied as a single use sterile 20 mg/ml solution for intravenous injection in 8.06 mM sodium phosphate, 1.47 mM potassium phosphate, 136.9 mM sodium chloride, 2.68 mM potassium chloride, and 0.01% polysorbate 80. 3BNC117 vials were shipped and stored at 4 °C.

Measurement of 3BNC117 serum levels. Serum levels of 3BNC117 were determined by a validated sandwich ELISA at Celldex Theapeutics as described previously[9]. Plates (Sigma-Aldrich PN: CLS3590 96-well, High Bind, polystyrene) were coated with 4 μg/ml of an anti-idiotypic antibody specifically recognizing 3BNC117 (anti-ID 1F1 mAb), and incubated overnight at 2–8 °C. After washing, plates were blocked for 1 h with 5% BSA. Serum samples, QCs and standards were added (1:50 minimum dilution in 5% BSA) and incubated for 1 h at room temperature. 3BNC117 was detected using an HRP-conjugated mouse anti-human IgG kappa chain specific antibody (Abcam PN: ab79115) and the HRP substrate tetra-methylbenzidine. 3BNC117 concentrations were then interpolated from a standard curve of 3BNC117 using a four-parameter logistic curve-fitting algorithm. The reference standard and positive controls were created from the drug product lot of 3BNC117 used in the clinical study.

Pharmacokinetic analysis. Blood samples were collected immediately before and at the end of infusions as well as on the day after infusion, weekly during the ATI period and at weeks 14, 24 and 36. 3BNC117 serum levels were measured by ELISA (Celldex Therapeutics).

Neutralization assay. Serum samples, viral supernatants, and control antibodies were tested against HIV-1 envelope pseudoviruses as previously described[36,37].

Cell-associated HIV-1 DNA. Participant's $CD4^+$ T-cells were isolated from 10 million cryopreserved PBMCs by negative magnetic selection (Miltenyi). Total DNA was extracted and quantitative PCR performed using *pol*- and CCR5-directed primers as previously described[7].

Virus cultures. Autologous virus was retrieved from HIV-1 infected individuals as previously described[31]. Briefly, healthy donor PBMCs were obtained by leukapheresis from a single donor. Cells were cultured at a concentration of 5×10^6/ml in Iscove's Modified Dulbecco's Medium (IMDM; Gibco) supplemented with 10% fetal bovine serum (FBS; HyClone, Thermo Scientific), 1% penicillin/streptomycin (Gibco), and 1 μg/ml phytohemagglutinin (PHA; Life Technologies) at 37 °C and 5% CO_2. After 2–3 days, 5×10^6 $CD8^+$ depleted cells were transferred into IMDM supplemented with 10% FBS, 1% penicillin/streptomycin, 5 μg/ml polybrene (Sigma), and 100 U/ml of IL-2. Cells were then co-incubated with $4–8 \times 10^6$ $CD4^+$ T cells from the study participants and 10 million irradiated healthy donor PBMCs that had been cultured together for 24 h prior in IMDM supplemented with 10% FBS, 1% penicillin/streptomycin, 100 U/ml IL-2 and 1 μg/ml PHA at 37 °C and 5% CO_2. Lymphoblasts were replenished weekly by adding 3 million healthy donor PHA stimulated $CD8^+$ depleted lymphoblasts. Culture supernatants were quantified using the Alliance HIV-1 p24 Antigen ELISA kit (PerkinElmer) according to the manufacturer's instructions. $TCID_{50}$s were determined for all HIV-1 containing supernatants[36,37] and then tested for sensitivity against 3BNC117 and other bNAbs in a TZM.bl neutralization assay. Blood samples and leukapheresis were collected under separate IRB-approved protocols and after volunteers provided informed consent.

Sequence analysis. HIV-1 RNA extraction and single genome amplification was performed as described previously[38]. In detail, HIV-1 RNA was extracted from plasma samples using the Qiagen MinElute Virus Spin kit (Qiagen) followed by first strand cDNA synthesis using SuperScript III reverse

transcriptase (Invitrogen Life Technologies) and the antisense primer env3out 5′-TTGCTACTTGTGATTGCTCCATGT-3′. gp160 *env* was amplified using envB5out 5′-TAGAGCCCTGGAAGCATCCAGGAAG-3′ and envB3out 5′-TTGCTACTTGTGATTGCTCCATGT-3′ in the first round and second round nested primers envB5in 5′-CACCTTAGGCATCTCCTATGGCAGGAAGAAG-3′ and envB3in 5′- GTCTCGAGATACTGCTCCCACCC-3′. PCRs were performed using a High Fidelity Platinum Taq (Invitrogen) at 94 °C, 2 min; (94 °C, 15 s; 55 °C 30 s; 68 °C, 4 min) × 35; 68 °C, 15 min. Second round PCR was performed with 1 μl of first PCR product as template and High Fidelity Platinum Taq at 94 °C, 2 min; (94 °C, 15 s; 55 °C 30 s; 68 °C, 4 min) × 35; 68 °C, 15 min. Sequence alignments, phylogenetic trees and mutation analysis of gp160 was performed by using Geneious Pro software, version 8.1.6 (Biomatters Ltd.)[39]. Sequence analysis was performed using Antibody database by A. West[30]. Logograms were generated using the Weblogo 3.0 tool[40].

Pseudovirus generation. Selected SGS from virus culture supernatants and plasma were used to generate pseudoviruses and tested for sensitivity to bNAbs in a TZM.bl assay[41]. To produce the pseudoviruses, plasmid DNA containing the cytomegalovirus (CMV) promoter was amplified by PCR using forward primer 5′-GTTGACATTGATTATTGACTAG and reverse primer 5′-CTTCCTGCCAT AGGAGATGCCTAAAGCTCTGCTTATATAGAC-CTC. The CMV promoter amplicon was fused to individual *env* SGS amplicons by PCR using forward primer 5′-AGTAATCAATTACGGGGTCATTAGTTCAT and reverse primer 5′-ACTTTTTGACCACTTGCCACCCAT. Fusion PCR was carried out using the Expand Long Template PCR System (Roche) in a 60 μl reaction consisting of 1 ng purified CMV promoter amplicon, 0.125 μl unpurified *env* SGA amplicon, 200 nM forward and reverse primers, 200 μM dNTP mix, 1× Buffer 1, and 1 μl DNA polymerase mix. PCR was run at 94 °C for 2 min; 25 cycles (94 °C for 12 s, 55 °C for 30 s, 68 °C for four minutes); and 72 °C for 10 min. Resulting amplicons were analysed by gel electrophoresis, purified without gel extraction, and co-transfected with pSG3Δ*env* into HEK293T cells to produce pseudoviruses as described previously[41].

Statistical analyses. Adverse events were summarized by the number of participants who experienced the event, by severity grade according to the DAIDS AE Grading Table and by relationship to 3BNC117 as determined by the investigator. PK-parameters were estimated by performing a non-compartmental analysis (NCA) using WinNonlin 6.3. Kaplan–Meier survival curves were used to compare time to rebound in trial participants to participants in previous ATI studies conducted by ACTG[29]. To exclude the possibility that the observed delay in rebound is confounded by clinical factors, we compared the clinical variables between the control (ACTG trial participants) and treated group using a two-sided Fisher's Exact test for categorical variables (gender and CD4 Nadir) and an unpaired Wilcoxon test (two-sided) for continuous variables (age, years on ART and CD4 count before ATI initiation) (Supplementary Table 7). Additionally, we tested for each potential confounder whether the variable is predictive for the rebound time. Therefore, we built a univariate survival regression model for each potential confounder and compared those models to a null model using a likelihood ratio test (LRT), which determines how much better the more complex model explains the data than the less complex model. Confounders were considered significant if the model with the potential confounder had an LRT P value of 0.05 or less, which was the case for 'years on ART' as well as 'age' for the comparison between the controls and the combined treatment group (Supplementary Table 7). We did not perform a standard Cox regression, since the proportional hazards assumption was not fulfilled for some of the variables. Rebound time was modelled using a log-normal distribution, which resulted in the best model fit as measured by Akaike information criterion (AIC) among several different distributions (Extended Data Fig. 4b–d). To determine the effect of the treatment after adjusting for the discovered confounders, we performed a weighted log-rank test[42]. Therefore, for each sample inverse probability weights based on the discovered confounders were estimated, which were used to re-weigh the variables of the log-rank statistic. We performed a bootstrapped version of the weighted log-rank test, as recommended in ref. 42 owing to the small sample size. We estimated the class probabilities using a lasso logistic regression model trained with the Matlab function lassoglm with five lambda values in a threefold cross-validation. To improve stability, the optimal lambda for the lasso logistic regression was determined only once using the original labels and used in all bootstrap runs to train the models that estimate the class probabilities.

Additionally, we performed an LRT at significance level $\alpha = 0.05$ based on a parametric survival regression model adjusted for the discovered confounders. In this analysis the treatment group still significantly predicted the delay in rebound (Supplementary Table 7). For the analyses, the R (version 3.2.1) packages survival (version 2.38-3) and fitdistrplus (version 1.0-6) were used and Matlab (version R2015b) for implementation of the weighted log-rank test.

Sequence and phylogenetic analysis. Nucleotide alignments were generated using ClustalW (v.2.11)[43] and manually adjusted using Geneious R8 (v.8.1.6)[39] and MacClade (v.4.08a)[44]. Sites that could not be unambiguously aligned were removed for all phylogenetic analyses. Optimal evolutionary model classes were determined using jModelTest (v.2.1.4)[45]. Maximum likelihood phylogenetic trees were generated using PhyML (v.3)[46] with joint estimation of evolutionary model parameter values and phylogenies. The tree comparing all participants was midpoint rooted and each within-subject tree was rooted on the basal branch as determined by the between-subject tree. Sequences with premature stop codons and frameshift mutations that fell in the gp120 surface glycoprotein region were excluded from all deduced protein analyses.

Sequences generated from the supernatants of viral outgrowth assays represented viruses that were present in the latent reservoir. Per assay, 4–8 million $CD4^+$ T cells were activated. In an HIV-infected person who is completely suppressed on antiretroviral therapy, it has been determined that 1×10^{-6} resting CD4 cells are latently infected with replication-competent virus[47]. Thus, one would expect to identify up to eight distinct viral isolates per individual culture. Single genome sequencing of the culture supernatants revealed sets of clonally related sequences, which appear as 'rakes' in a phylogenetic tree (Extended Data Figs 7 and 8). The most recent common ancestor of these rakes represents the reactivated virus that was present in the host (similar to the inference of infectious molecular clones as described in ref. 48). As shown in Extended Data Figs 7 and 8, sequences from culture reactions fall in 1–3 rakes within a given individual. We inferred each rake's most recent common ancestor (MRCA) by building a majority-rule consensus and treated it as a single virus from the participant's latent pool. These MRCAs were used to build the phylogenetic trees shown in Fig. 3.

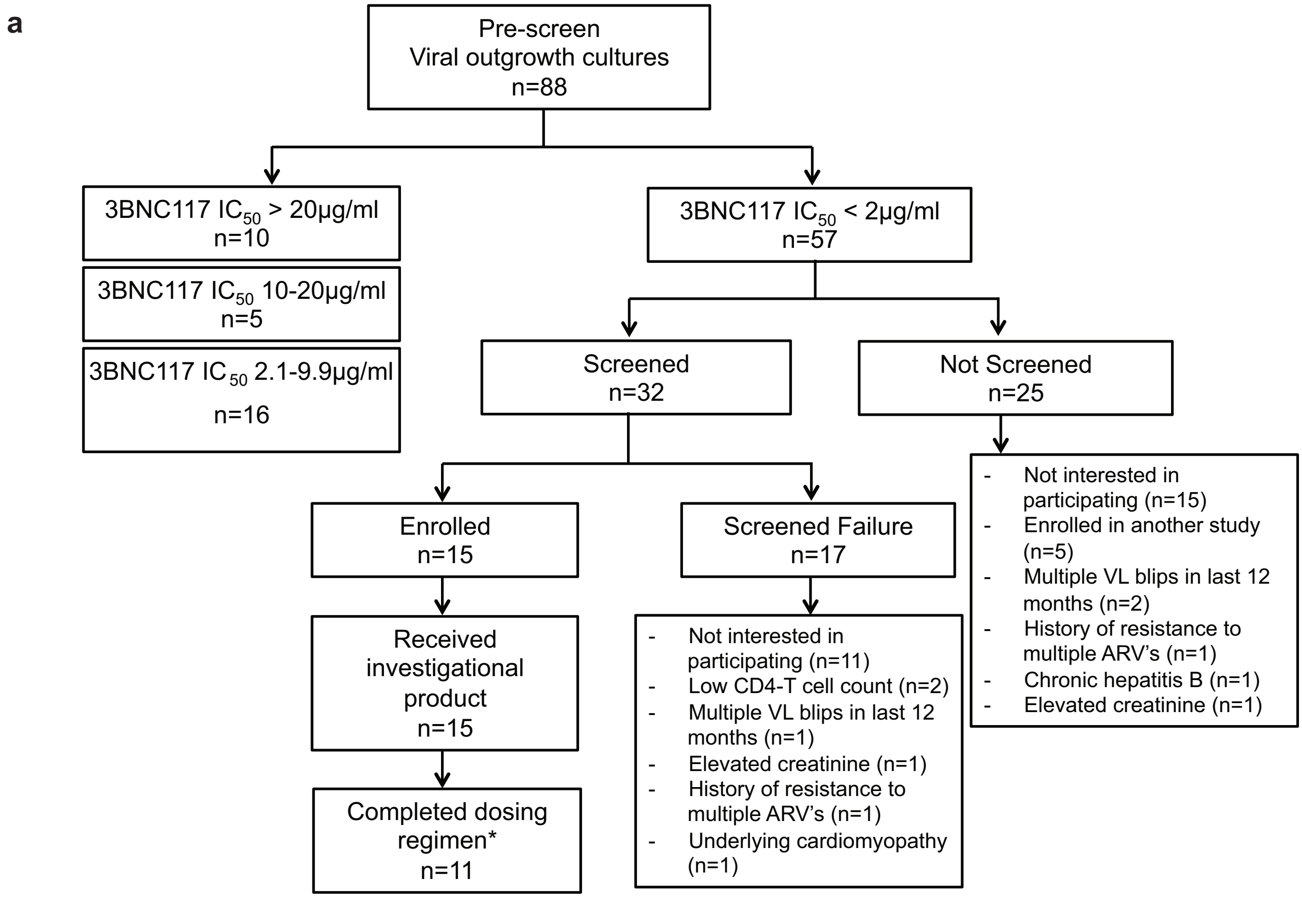
Because mixed culture isolates replicated for 14 or more days, *in vitro* recombinants were observed. *In vitro* recombinants from culture reactions were identified and removed from the data set if they: (i) had two identifiable parental sequences within the same culture reaction; and (ii) exhibited three consecutive informative sites relative to one parent followed by three consecutive informative sites relative to another. We independently verified that a subset of these sequences showed evidence of recombination using the Recco tool (v.0.93)[49].

Assessment of rebound virus clonality. The Poisson Fitter v2 tool[50] is designed to test if a set of homogeneous sequences exhibits random diversification. If such a set of sequences exhibits a star-like phylogeny and a Poisson distribution of pairwise differences (Hamming distances), it can be inferred that a single virus gave rise to those sequences. Poisson Fitter v2 tests these and other parameters using maximum likelihood methods and performs a χ^2 goodness of fit test to obtain a P value. A non-significant P value signifies that the observed Hamming distances adhere to a Poisson distribution and it can be inferred that a single virus gave rise to rebound. Single genome derived *env* sequences from the plasma at the earliest time point post-rebound from each participant were tested using Poisson Fitter (Supplementary Table 8).

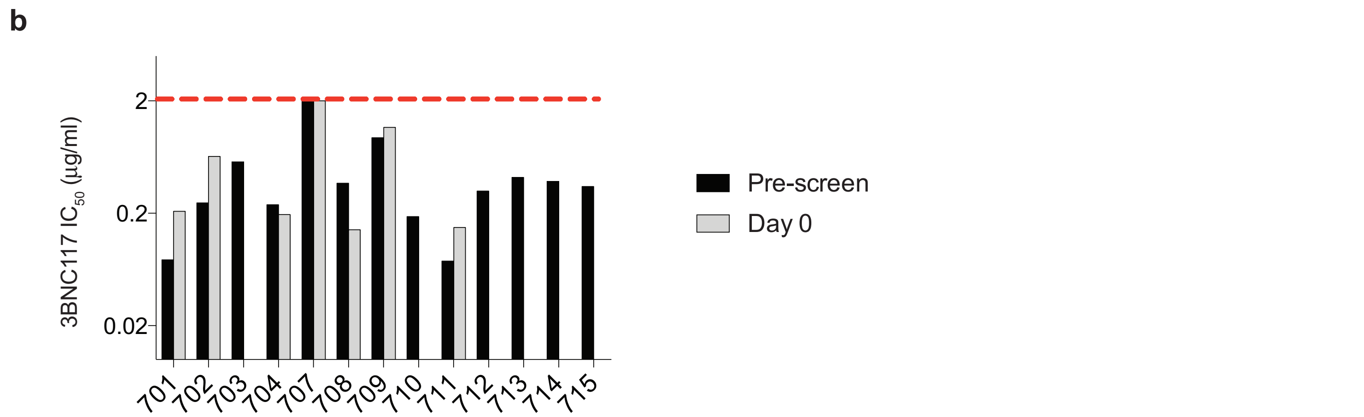
ELISPOT T-cell response analysis. Interferon gamma Elispot assays were performed as described[51]. Briefly, 96-well polyvinylidene plates (Millipore, Bedford, Mass.) were precoated with 0.5 g/ml of anti-IFNγ monoclonal antibody, 1-DIK (Mabtech, Stockholm, Sweden) and previously frozen PBMCs were plated at a concentration of 50,000 to 100,000 cells per well in a volume of 100 μl of R10 medium (RPMI 1640 (Sigma), 10% fetal calf serum (Sigma), 10 mM HEPES buffer (Sigma)) with antibiotics (2 mM L-glutamine, 50 U of penicillin-streptomycin/ml). Plates were incubated overnight at 37 °C, 5% CO_2, and developed as described[51]. Cells were tested against a panel of 410 B-clade overlapping 18-mer peptides (OLPs) spanning the entire HIV-1 genome (consensus sequence from 2001). These peptides were used in a matrix system of 11–12 peptides per pool to screen study participants for HIV-specific T cell responses. Confirmation of recognized individual peptides within a peptide pool was undertaken in an additional Elispot assay, as described[52]. Wells containing PBMCs and R10 medium alone were used as negative controls and were run in duplicate on each plate. Wells containing PBMCs and phytohemagglutinin (PHA) served as positive controls. The numbers of spots per well were counted using an automated Elispot plate reader (ImmunoSpot Reader System; Cellular Technology Limited, Shaker Heights, OH). Responses were regarded as positive if they had at least three times the mean number of spot forming cells (SFC) in the two negative control wells and had to be >50 SFC/10^6 PBMCs (ref. 51,52).

31. Laird, G. M. *et al.* Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* **9,** e1003398 (2013).
32. Volberding, P. *et al.* Antiretroviral therapy in acute and recent HIV infection: a prospective multicenter stratified trial of intentionally interrupted treatment. *AIDS* **23,** 1987–1995 (2009).
33. Kilby, J. M. *et al.* A randomized, partially blinded phase 2 trial of antiretroviral therapy, HIV-specific immunizations, and interleukin-2 cycles to promote efficient control of viral replication (ACTG A5024). *J. Infect. Dis.* **194,** 1672–1676 (2006).

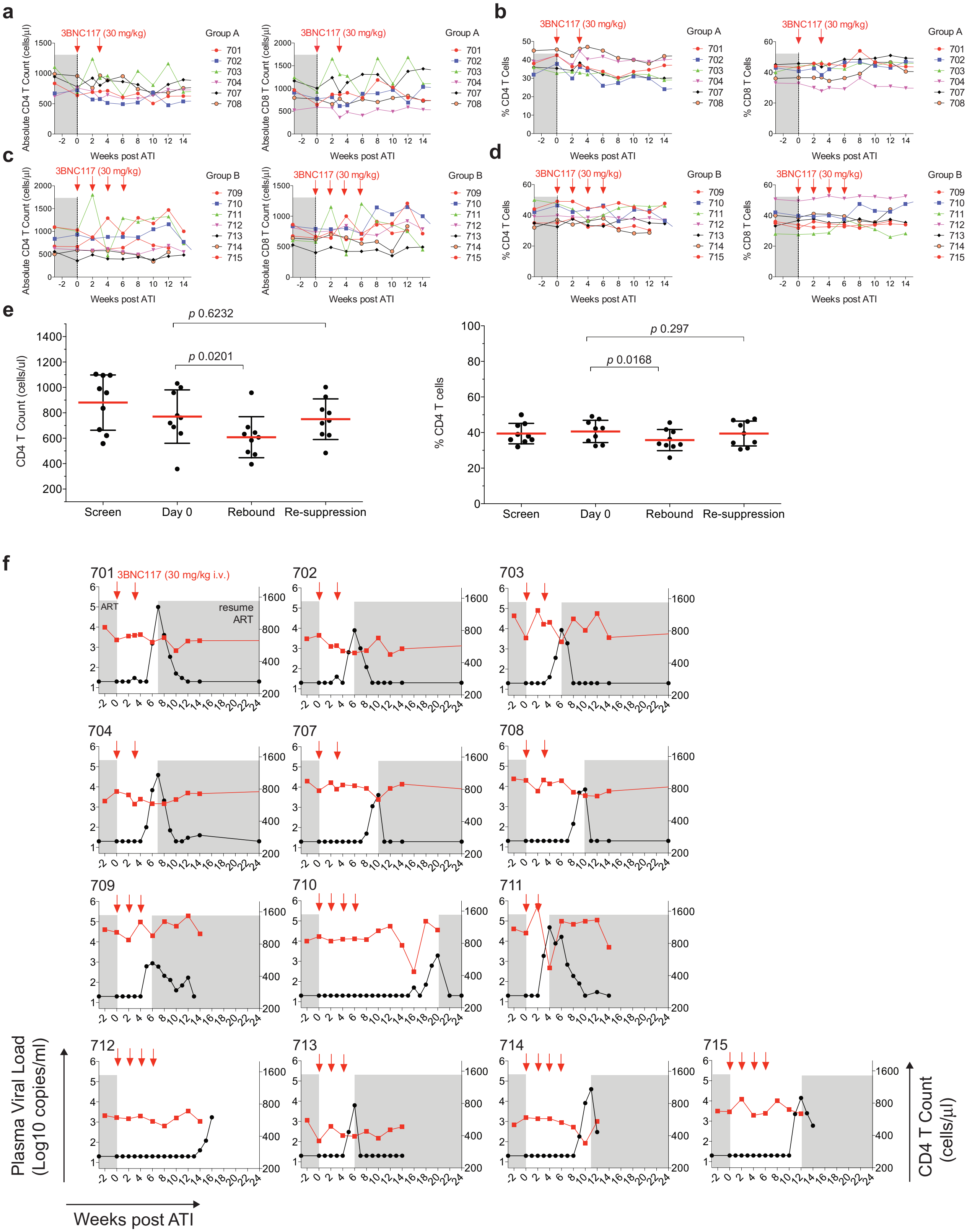
34. Jacobson, J. M. *et al.* Evidence that intermittent structured treatment interruption, but not immunization with ALVAC-HIV vCP1452, promotes host control of HIV replication: the results of AIDS Clinical Trials Group 5068. *J. Infect. Dis.* **194,** 623–632 (2006).
35. Pereyra, F. *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330,** 1551–1557 (2010).
36. Montefiori, D. C. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr. Protoc. Immunol.* **Chapter 12**, Unit 12 11 (2005).
37. Li, M. *et al.* Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* **79,** 10108–10125 (2005).
38. Salazar-Gonzalez, J. F. *et al.* Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J. Virol.* **82,** 3952–3970 (2008).
39. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).
40. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
41. Kirchherr, J. L. *et al.* High throughput functional analysis of HIV-1 env genes without cloning. *J. Virol. Methods* **143,** 104–111 (2007).
42. Xie, J. & Liu, C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat. Med.* **24,** 3089–3110 (2005).
43. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).
44. Maddison, W. P. & Maddison, D. R. *MacClade – Analysis of Phylogeny and Character Evolution – Version 4.* (Sinauer Associates, Inc., 2001).
45. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772 (2012).
46. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).
47. Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387,** 183–188 (1997).
48. Parrish, N. F. *et al.* Phenotypic properties of transmitted founder HIV-1. *Proc. Natl Acad. Sci. USA* **110,** 6626–6633 (2013).
49. Maydt, J. & Lengauer, T. Recco: recombination analysis using cost optimization. *Bioinformatics* **22,** 1064–1071 (2006).
50. Giorgi, E. E. & Bhattacharya, T. A note on two-sample tests for comparing intra-individual genetic sequence diversity between populations. *Biometrics* **68,** 1323–1326, author reply 1326 (2012).
51. Altfeld, M. A. *et al.* Identification of dominant optimal HLA-B60- and HLA-B61-restricted cytotoxic T-lymphocyte (CTL) epitopes: rapid characterization of CTL responses by enzyme-linked immunospot assay. *J. Virol.* **74,** 8541–8549 (2000).
52. Addo, M. M. *et al.* Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J. Virol.* **77,** 2081–2092 (2003).



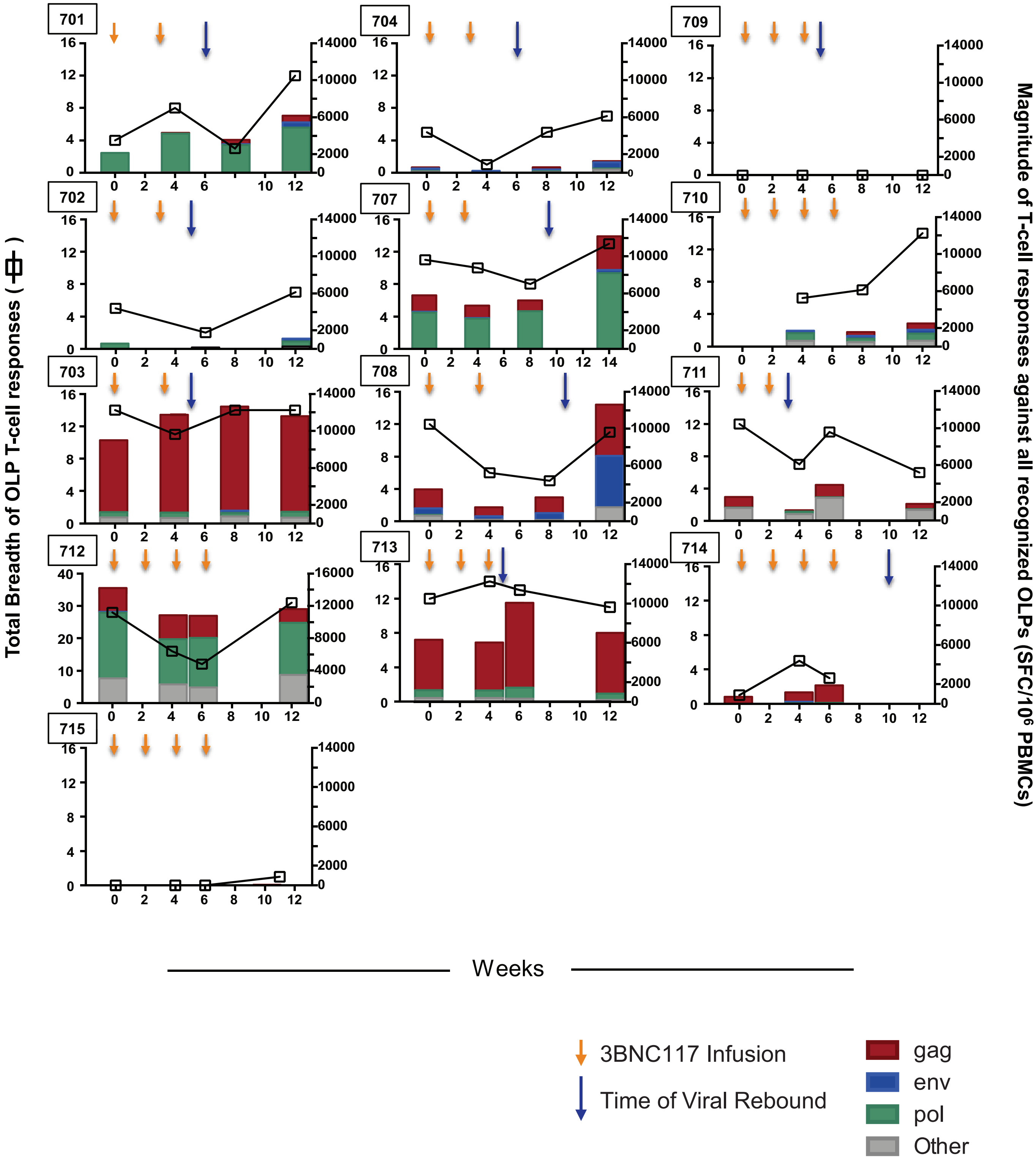
*Participant 705 received only 1 dose due to VL > 1,000 at day 0. Some participants experienced viral rebound prior to completion of all 4 scheduled infusions: 709 – received 3 doses; 711 - 2 doses; 713 - 3 doses.



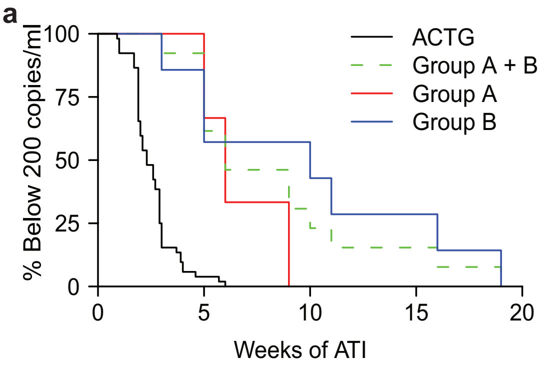
Extended Data Figure 1 | Study participant selection and neutralization of pre-infusion cultures by 3BNC117. a, Flow diagram showing the selection of study participants. **b**, Bar diagrams showing IC_{50} values ($\mu g\ ml^{-1}$) in TZM-bl assays for 3BNC117 against bulk virus outgrowth culture supernatants from the indicated time point pre-infusion for each participant (Supplementary Table 3). For some participants both screen and day 0 cultures were obtained and showed less than threefold variation in IC_{50} values. The red dotted line indicates an IC_{50} of $2\ \mu g\ ml^{-1}$ which was used as a threshold for inclusion in the study.

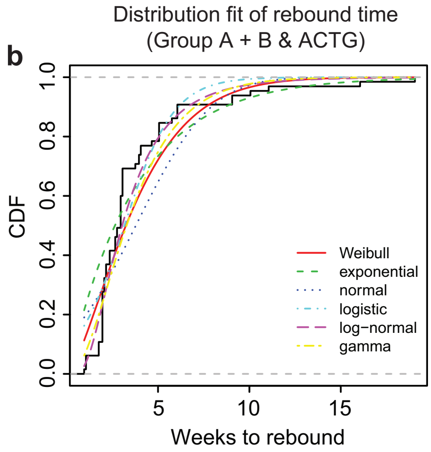


Extended Data Figure 2 | CD4$^+$ and CD8$^+$ T cells during study period in participants. a–d, Absolute T cell counts (**a**, **c**) and percentage of CD4$^+$ and CD8$^+$ T cells among CD3$^+$ T cells (**b**, **d**) for group A and B, respectively (Supplementary Table 4). 3BNC117 infusions are indicated with red arrows. **e**, Comparison of absolute CD4$^+$ T cell counts and percentage of CD4$^+$ T cells among CD3$^+$ T cells at screen, day 0, rebound and after re-suppression. Shown is the data from participants 701, 702, 703, 704, 707, 708, 709, 711 and 713 for whom re-suppression CD4 counts were available (Supplementary Table 4). The last available time point was used as re-suppression time point. Red lines indicate the mean value and error bars indicate standard deviation. *P* values were obtained using a paired *t*-test comparing the indicated time points. **f**, Plasma viral loads and CD4 counts in all study participants. 3BNC117 infusions are indicated with red arrows. The left *y*-axis shows plasma viral loads in RNA copies per ml (black curves), and right *y*-axis shows absolute CD4 counts in cells per μl (red curves). Grey areas indicate ART therapy.

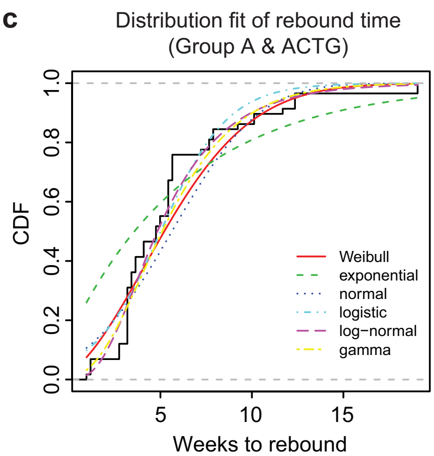


Extended Data Figure 3 | HIV-specific T-cell responses. Total breadth (open squares) and magnitude (bars) of T-cell responses against HIV-1 overlapping peptides (OLPs) at the designated time points following administration of 3BNC117 (yellow arrows indicate infusions of 3BNC117 at 30 mg kg^{-1}). For all study participants, antiretroviral therapy was discontinued on day 2 after the first 3BNC117 administration. Blue arrows indicate the time of viral rebound. For study participants 710, 712 and 715 rebound occurred at week 19, 16 and 11, respectively. Baseline samples for study participant 710 and week 12 samples for study participant 714 were not available for ELISpot analysis. Overall, breadth, magnitude and protein specificity were heterogeneous among the study participants.

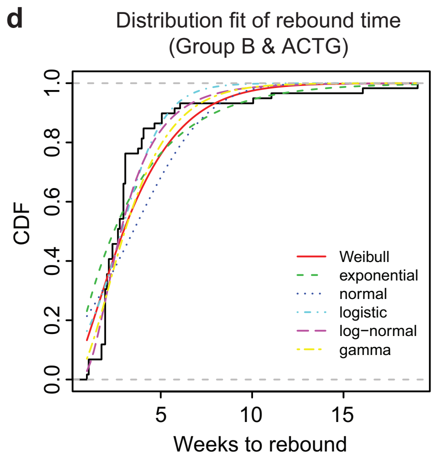




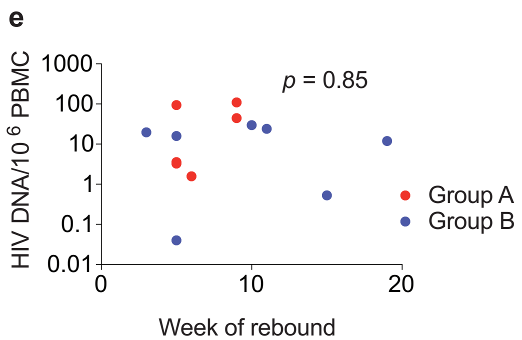
Distribution	AIC	BIC
Weibull	292.74	297.09
exponential	303.11	305.28
normal	340.57	344.91
logistic	314.41	318.76
log-normal	267.44	271.79
gamma	283.06	287.41



Distribution	AIC	BIC
Weibull	212.85	216.97
exponential	245.04	247.10
normal	228.66	232.78
logistic	218.50	222.62
log-normal	197.98	202.10
gamma	203.48	207.60

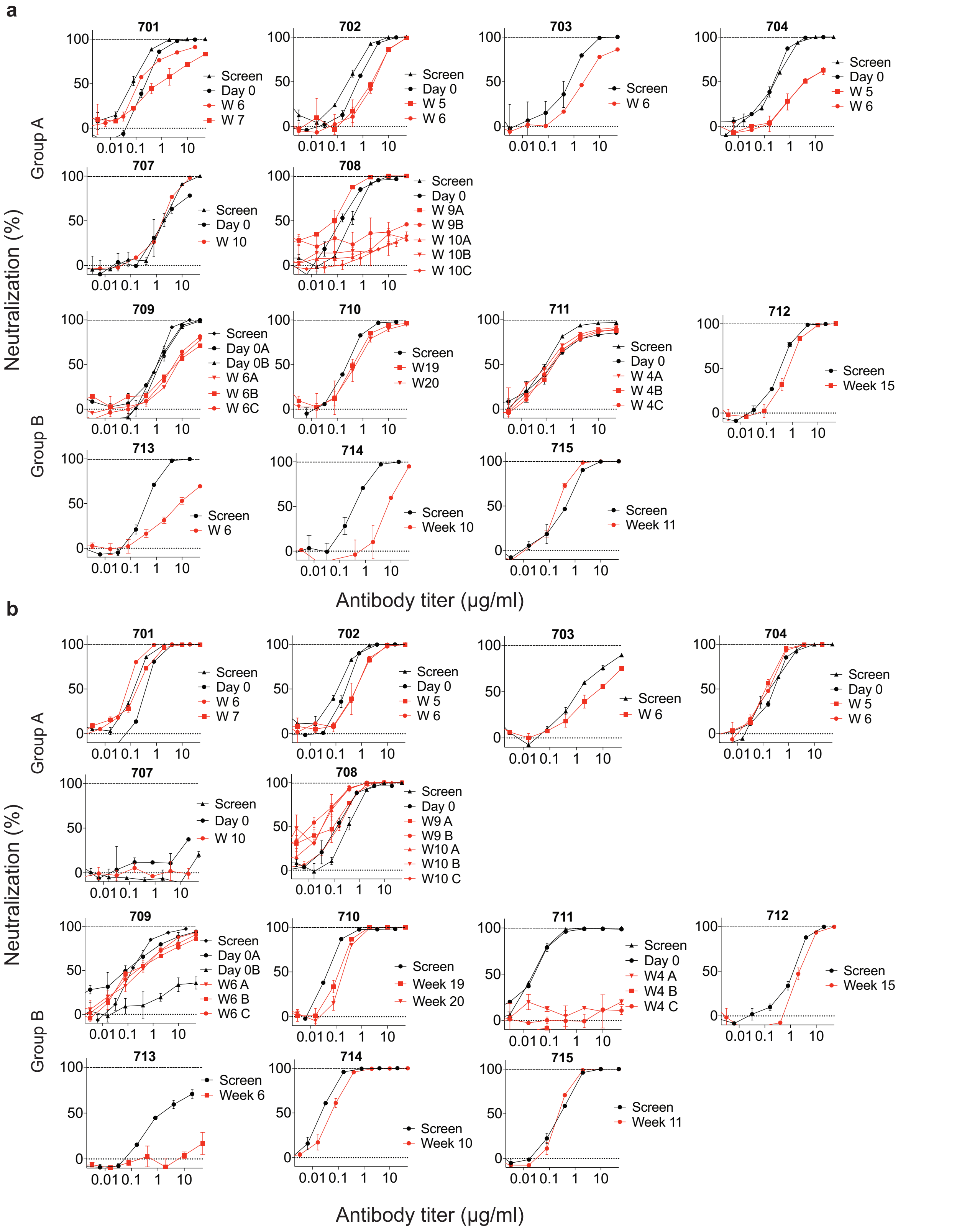


Distribution	AIC	BIC
Weibull	257.36	261.52
exponential	265.46	267.54
normal	308.32	312.48
logistic	273.96	278.12
log-normal	228.17	232.33
gamma	246.67	250.82

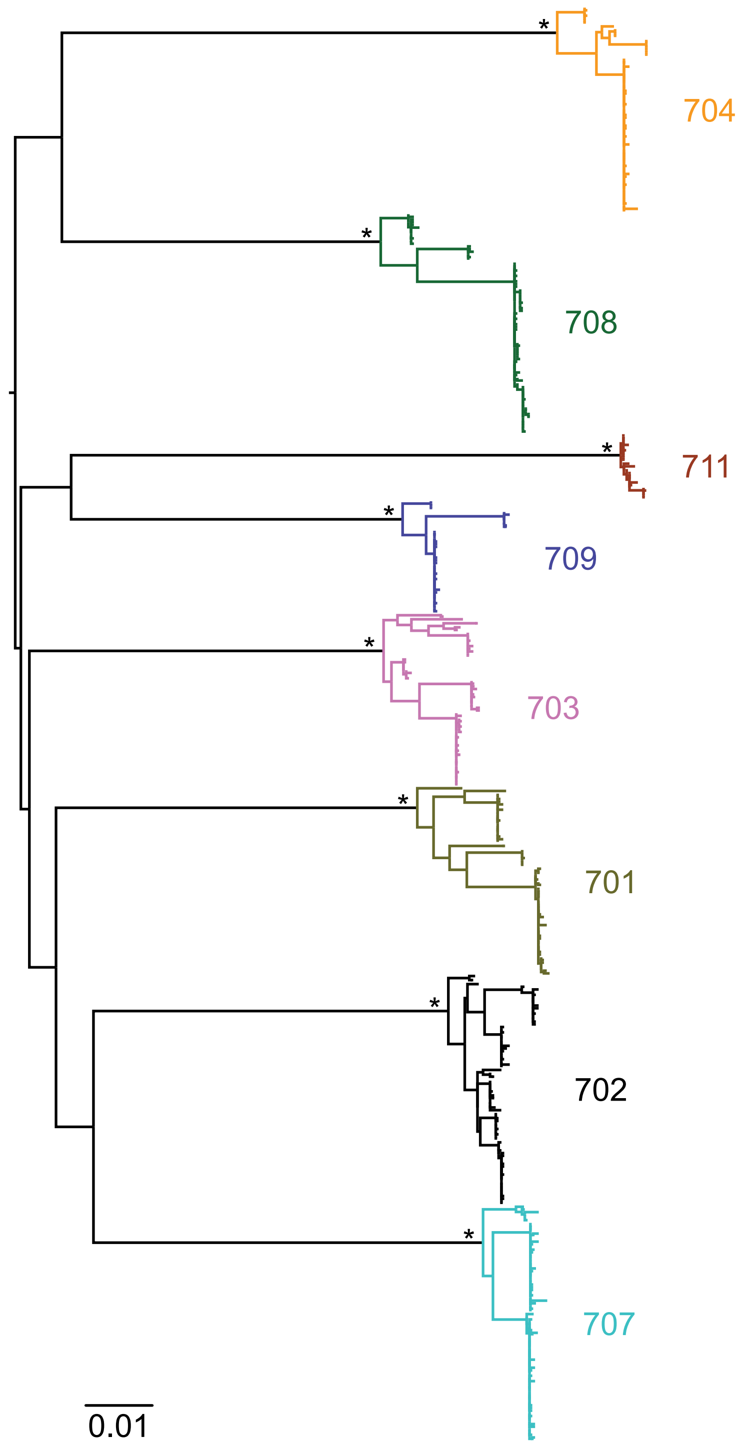


Extended Data Figure 4 | See next page for caption.

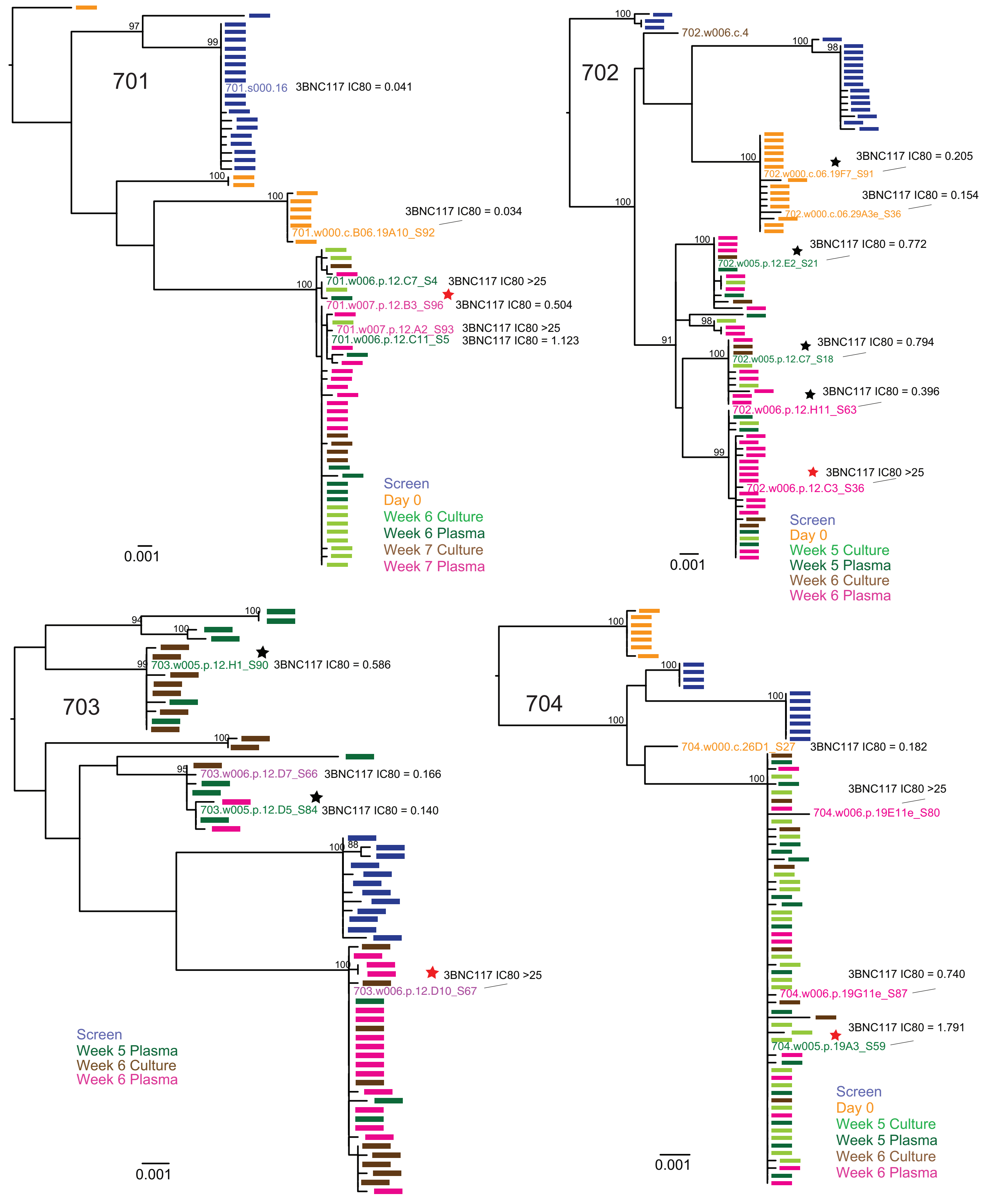
Extended Data Figure 4 | Viral rebound in ACTG control subjects and trial participants. a, Kaplan–Meier plot summarizing viral rebound in 52 ACTG trial participants who underwent ATI without antibody treatment (black, Supplementary Table 6) and trial participants (Fig. 2a, b, Supplementary Table 4). Six group A participants are shown in red, seven Group B participants in blue and the combination in green as indicated. The y-axis indicates the percentage of participants with viral levels below 200 RNA copies per ml, x-axis indicates weeks after ATI initiation. The survival curves of all considered partitions of the trial participants (group A, group B and group A + B) differed significantly at significance level $\alpha = 0.05$ from the survival curve of the ACTG trial participants. For the comparison of group A (group A + B) with the ACTG trial participants, we performed a weighted log-rank test adjusting for the clinical variables 'years on ART' and 'age' to correct for possible confounding factors (Supplementary Table 7, $P < 0.00001$). We identified those potential confounders by univariate parametric survival regression using a likelihood ratio test (Statistical Methods). Since we did not discover any confounders with the same analysis among all available clinical variables for the comparison between group B participants and the ACTG trial participants, we performed a standard log-rank test in that setting ($P < 0.0001$). **b–d**, In order to perform a survival regression, the distribution of the rebound times has to be determined. Therefore, we compared the empirical cumulative distribution function (CDF) of the rebound times (black, solid line) with the CDF of the rebound times to a fitted distribution (Weibull, exponential, normal, logistic, log-normal, and gamma) for each comparison group (combined trial participants, group A or group B with ACTG control patients). Since the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were smallest for the log-normal distribution (green), we have chosen to model the rebound times with the log-normal distribution. **e**, Dot plot indicating the relationship between cell associated HIV DNA in pre-infusion PBMCs (y-axis) and the week of rebound (x-axis). Group A and group B participants are coloured red and blue respectively. The P value was derived from calculating the Pearson correlation coefficient.



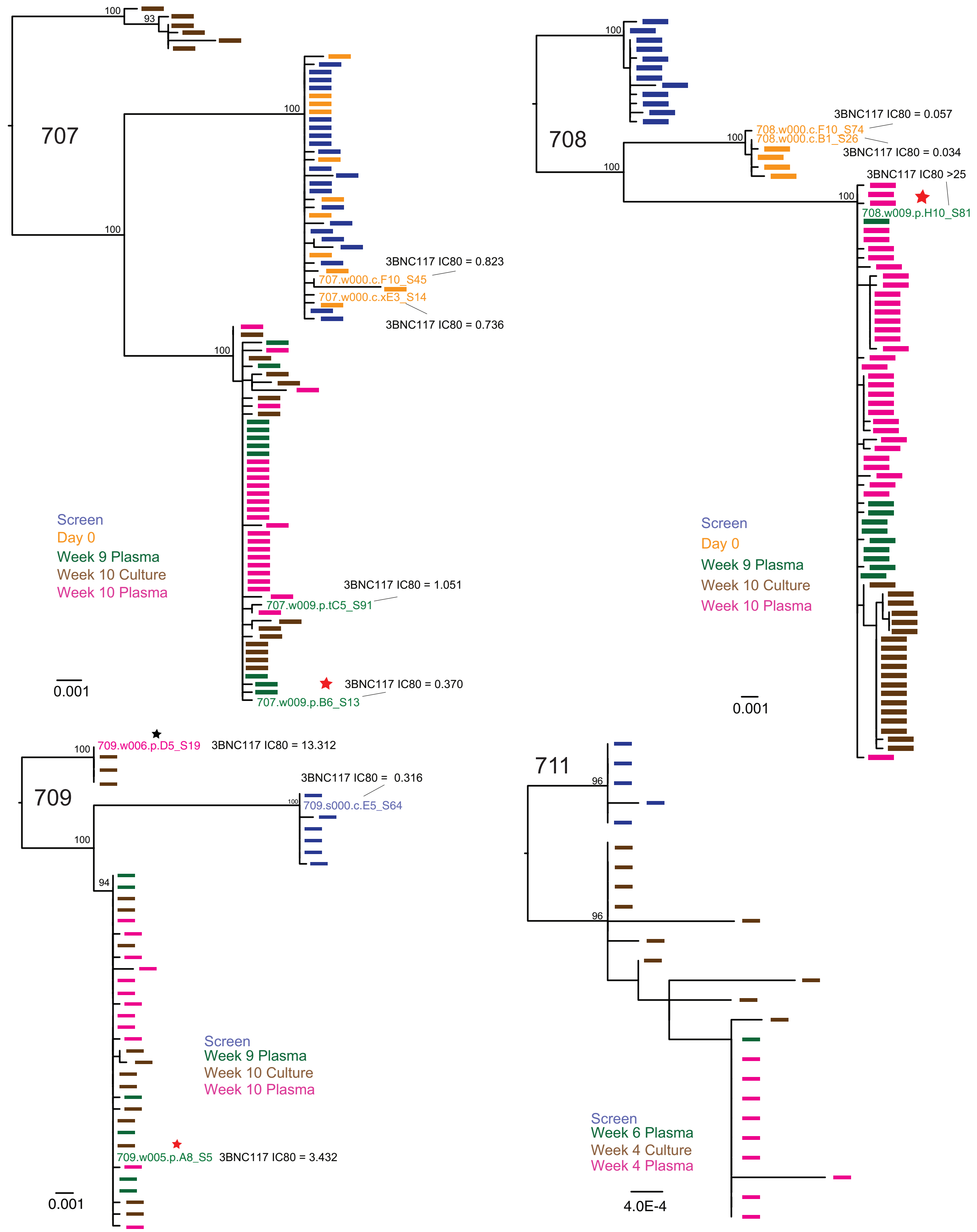
Extended Data Figure 5 | *In vitro* neutralization of pre-infusion and rebound virus outgrowth cultures by 3BNC117 or 10-1074. a, b, TZM-bl assay neutralization by 3BNC117 (**a**) and 10-1074 (**b**) are shown for individual virus outgrowth cultures derived from pre-infusion (black lines/symbols) or rebound (red lines/symbols) time points for each participant. In some cases, multiple independent cultures were grown from a single time point and assayed for neutralization (Supplementary Table 3). 'Screen' refers to cultures of PBMC samples taken weeks before infusion during screening, while 'Day 0' refers to cultures of PBMCs collected immediately before the first 3BNC117 infusion. Rebound culture time points are denoted by the week (W) at which the samples were collected. Symbols reflect the means of two technical replicates; error bars denote standard deviation.



Extended Data Figure 6 | Phylogenetic tree of *env* nucleotide sequences from trial participants. A maximum likelihood phylogenetic tree was constructed from single-genome-derived viral *env* sequences from outgrowth culture supernatants as well as plasma from participants 701 (olive), 702 (black), 703 (pink), 704 (yellow), 707 (light blue), 708 (green), 709 (dark blue) and 711 (brown). Hypervariable (as defined in http://www.hiv.lanl.gov/content/sequence/VAR_REG_CHAR/) and other poorly aligned regions were excluded from the analysis. The tree was constructed using PhyML with a GTR+I+G substitution model and midpoint rooted. Asterisks indicate 100% bootstrap support (only values for major nodes are shown). The scale bar indicates 0.01 substitutions per site.



Extended Data Figure 7 | Rebound virus clonality and neutralization sensitivity to 3BNC117. Maximum likelihood phylogenetic trees of plasma and culture-derived *env* sequences are shown for participants 701, 702, 703, 704. Sequences obtained at screening, on Day 0, and consecutive rebound time points (plasma and cultures) are colour coded as indicated. The trees were rooted based on the branch insertion identified in the between-subject tree (Extended Data Fig. 6). Bootstrap values ≥90% are shown. Names of *env* sequences used to generate pseudoviruses for 3BNC117 neutralization analysis are indicated along with the respective IC_{80} titres in μg ml^{-1}. Representative rebound viruses selected in Fig. 4b are marked with red stars (Fig. 4b, Supplementary Table 9). Zero branch length viruses in multi-rebounders 702 and 703 are marked with black stars.



Extended Data Figure 8 | Rebound virus clonality and neutralization sensitivity to 3BNC117. Maximum likelihood phylogenetic trees of plasma- and culture-derived *env* sequences are shown for participants 707, 708, 709 and 711. Sequences obtained at screening, on day 0, and consecutive rebound time points (plasma and cultures) are colour coded as indicated. The trees were rooted based on the branch insertion identified in the between-subject tree (Extended Data Fig. 6). Bootstrap values ≥90% are shown. Names of *env* sequences used to generate pseudoviruses for 3BNC117 neutralization analysis are indicated along with the respective IC_{80} titres in μg ml^{-1}. Representative rebound viruses selected in Fig. 4b are marked with red stars (Fig. 4b, Supplementary Table 9). Zero branch length virus in multi-rebounder 709 is marked with a black star.

doi:10.1038/nature18618

Mitochondrial and nuclear DNA matching shapes metabolism and healthy ageing

Ana Latorre-Pellicer[1,2], Raquel Moreno-Loshuertos[3], Ana Victoria Lechuga-Vieco[1,4], Fátima Sánchez-Cabo[1], Carlos Torroja[1], Rebeca Acín-Pérez[1], Enrique Calvo[1], Esther Aix[1], Andrés González-Guerra[1], Angela Logan[5], María Luisa Bernad-Miana[6], Eduardo Romanos[6], Raquel Cruz[2], Sara Cogliati[1], Beatriz Sobrino[7], Ángel Carracedo[2,7,8], Acisclo Pérez-Martos[3], Patricio Fernández-Silva[3], Jesús Ruíz-Cabello[1,4,9], Michael P. Murphy[5], Ignacio Flores[1], Jesús Vázquez[1] & José Antonio Enríquez[1,3]

Human mitochondrial DNA (mtDNA) shows extensive within-population sequence variability[1]. Many studies suggest that mtDNA variants may be associated with ageing or diseases[2–4], although mechanistic evidence at the molecular level is lacking[5,6]. Mitochondrial replacement has the potential to prevent transmission of disease-causing oocyte mtDNA. However, extension of this technology requires a comprehensive understanding of the physiological relevance of mtDNA sequence variability and its match with the nuclear-encoded mitochondrial genes. Studies in conplastic animals[7–9] allow comparison of individuals with the same nuclear genome but different mtDNA variants, and have provided both supporting and refuting evidence that mtDNA variation influences organismal physiology. However, most of these studies did not confirm the conplastic status, focused on younger animals, and did not investigate the full range of physiological and phenotypic variability likely to be influenced by mitochondria. Here we systematically characterized conplastic mice throughout their lifespan using transcriptomic, proteomic, metabolomic, biochemical, physiological and phenotyping studies. We show that mtDNA haplotype profoundly influences mitochondrial proteostasis and reactive oxygen species generation, insulin signalling, obesity, and ageing parameters including telomere shortening and mitochondrial dysfunction, resulting in profound differences in health longevity between conplastic strains.

The mtDNAs of C57BL/6 and NZB/OlaHsd mice differ by 12 missense mutations, 4 transfer RNA (tRNA) mutations, 8 ribosomal RNA (rRNA) mutations, and 10 non-coding-region mutations (Extended Data Fig. 1b, Extended Data Table 1 and Supplementary Data 1), a level of divergence comparable to that between human Eurasian and African mtDNAs. We developed a conplastic mouse strain (Extended Data Fig. 1a) with the C57BL/6 nuclear genome and the NZB/OlaHsd mtDNA, hereafter referred to as BL/6^{NZB}. The original C57BL/6 strain with the C57BL/6 nuclear genome and C57BL/6 mitochondrial genome is hereafter referred to as BL/6^{C57}. Nuclear genome purity was confirmed (Extended Data Fig. 1c–e). Conplastic animals, BL/6^{C57} and BL/6^{NZB}, were fertile, and on a chow diet there were no weight differences throughout the growth period (Extended Data Fig. 2a, b), with marginal differences in activity observed in metabolic chambers (Extended Data Fig. 2c–e).

Unexpectedly BL/6^{NZB} extended median lifespan by 16% without modifying maximum lifespan (Fig. 1a). After 9 months, BL/6^{NZB} mice gained less weight (Fig. 1b). Two-year-old BL/6^{C57} animals manifested more signs of ageing[10] than age-matched BL/6^{NZB} mice (Fig. 1c, Extended Data Fig. 3a, b); including ovary preservation, with less lipofuscin accumulation at 12 months in BL/6^{NZB} mice (Extended Data Fig. 3e, f). Telomeres in one-year-old animals were on average approximately 11% longer in BL/6^{C57} mice, whereas in two-year-old animals this was inverted, with BL/6^{NZB} telomeres slightly longer (Fig. 1d). This is due to a BL/6^{C57} telomere-length-reduction rate that is double than that of BL/6^{NZB} mice (Fig. 1e). Liver necropsy revealed a higher incidence of tumours in BL/6^{C57} mice at death (Fig. 1f and Extended Data Fig. 3c, d), suggesting greater genome instability.

Four days after birth, respiration and ATP synthesis were lower in liver mitochondria from BL/6^{NZB} mice. Both variables subsequently equalized, but from 100-days-old onwards, the rates in BL/6^{C57} liver mitochondria decreased progressively (Fig. 1g). These changes were due to a steady decline in BL/6^{C57} respiration, with BL/6^{NZB} respiration remaining constant between days 20 and 300 (Fig. 1h). Liver mtDNA sequencing revealed no age-related increase in mtDNA mutations in either mtDNA backgrounds (Extended Data Table 2, Supplementary Data1). Later metabolic deterioration, lower tumour incidence, slower telomere attrition[11], and sustained ovarian function indicate that the NZB mtDNA promotes healthier ageing in the BL/6 nuclear background.

Gene expression signatures of 12-week-old mice (Fig. 2a and Supplementary Data 2) revealed differences in molecular functions related to 'lipid and carbohydrate metabolism' in liver and heart and 'free radical scavenging' specifically in liver (Fig. 2b). Consistent with increased PPARα activity, BL/6^{NZB} mice had higher expression of lipid metabolism and lower expression of carbohydrate metabolism and inflammation pathways (Extended Data Fig. 4 and Supplementary Data 2).

Plasma and liver metabolomics of 12–week-old mice confirm the mtDNA haplotype effect in three main groups: glutathione metabolism; digestion and absorption; and lipid and membrane metabolism (Fig. 2c, Extended Data Fig. 5 and Supplementary Data 3). Metabolite set enrichment analysis (MSEA) of glutathione metabolism revealed enhanced glutathione turnover[12] and reactive oxygen species (ROS) defence in BL/6^{NZB} animals (Fig. 2d and Extended Data Fig. 5b). In parallel, we detected lower levels of the hepatic lipid peroxidation product 12,13-hydroxyoctadec-9(Z)-enoate. Other affected metabolites included a number of fatty acids and lysophospholipids, consistent with the differences in membrane metabolism (Extended Data Fig. 5d).

Quantitative proteomics revealed a differential profile for the ROS defence proteome (Fig. 2e, f and Supplementary Data 4). In liver, the major mitochondria-located redox regulators Glrx2, Glrx5 and Txnrd2 were increased in young and old BL/6^{NZB} mice, whereas

[1]Centro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid 28029, Spain. [2]Grupo de Medicina Xenómica, CIBERER, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain. [3]Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, Zaragoza 50009, Spain. [4]CIBERES, C/ Melchor Fernández-Almagro 3, 28029 Madrid, Spain. [5]Medical Research Council Mitochondrial Biology Unit, Hills Road, Cambridge CB2 0XY, UK. [6]SAI de Biomedicina y Biomateriales & Phenotypic Unit, Instituto Aragonés de Ciencias de la salud, Zaragoza 50009, Spain. [7]Fundación Pública Gallega Medicina Xenómica, SERGAS, Complexo Hospitalario Universitario de Santiago, 15706 Santiago de Compostela, Spain. [8]Center of Excellence in Genomic Medicine Research, King Abdulaziz University, 21589 Jeddah, Saudi Arabia. [9]Universidad Complutense de Madrid, Madrid 28606, Spain.

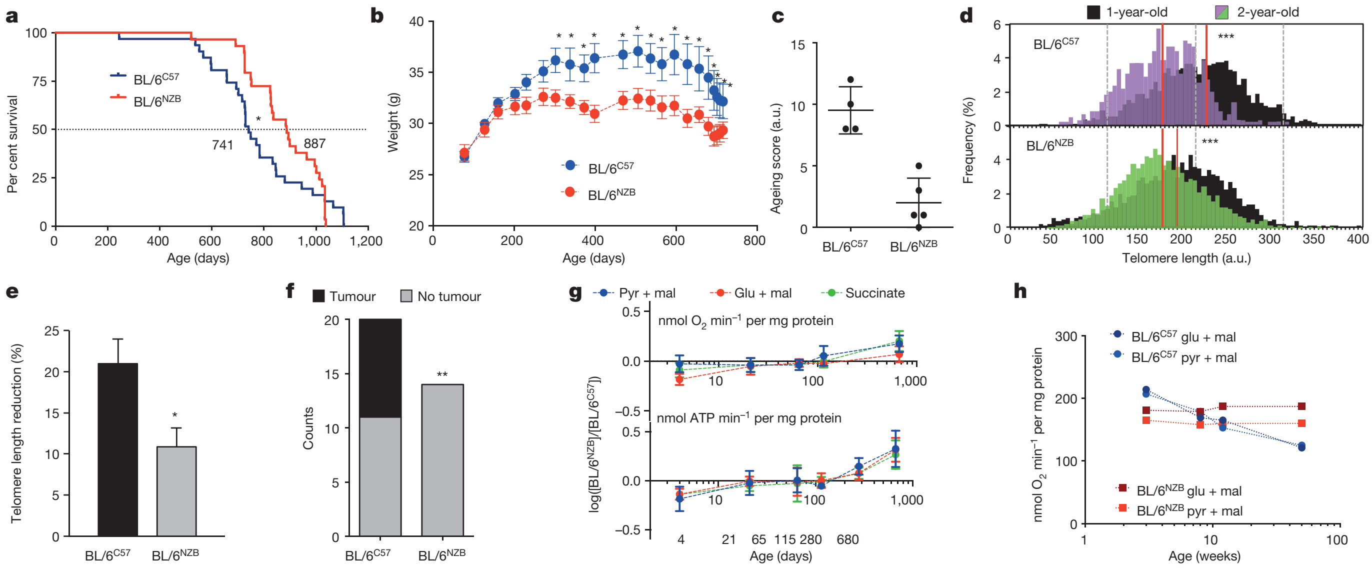


Figure 1 | Conplastic animals differ in metabolism and ageing profiles. **a**, Survival curves ($n = 31$ per genotype (19 males, 12 females); Gehan–Breslow–Wilcoxon test $*P < 0.05$). **b**, Weight gain (chow diet, $n = 19$ males per genotype, mean ± s.e.m., $*P < 0.05$, two-tailed t-test). **c**, Mice that were ~2 years old were scored for ageing parameters ($n = 5$ for BL/6^{NZB}, $n = 4$ for BL/6^{C57}). For more information on ageing parameters see Extended Data Fig. 3a. a.u., arbitrary units. **d**, Telomere length in hair follicle cells of 1- ($n = 6$ per group) and 2-year-old mice ($n = 5$ for BL/6^{NZB}, $n = 4$ for BL/6^{C57}). Red lines indicate mean length ($***P < 0.001$, Wilcoxon's rank sum test). **e**, Telomere length reduction during second year of life (mean ± s.e.m. $*P < 0.05$, two-tailed t-test). **f**, Tumour incidence in animals dying from natural causes ($n = 14$ for BL/6^{NZB}, $n = 20$ for BL/6^{C57}; $**P < 0.01$, Fisher's exact test). **g**, Relative O_2 consumption (coupled) and ATP synthesis rates in liver mitochondria ($n = 4$ per genotype and age; mean ± s.d.). **h**, O_2 consumption (coupled) respiration in liver mitochondria simultaneously isolated from 22-, 30-, 120- and 300-day-old males ($n = 8$).

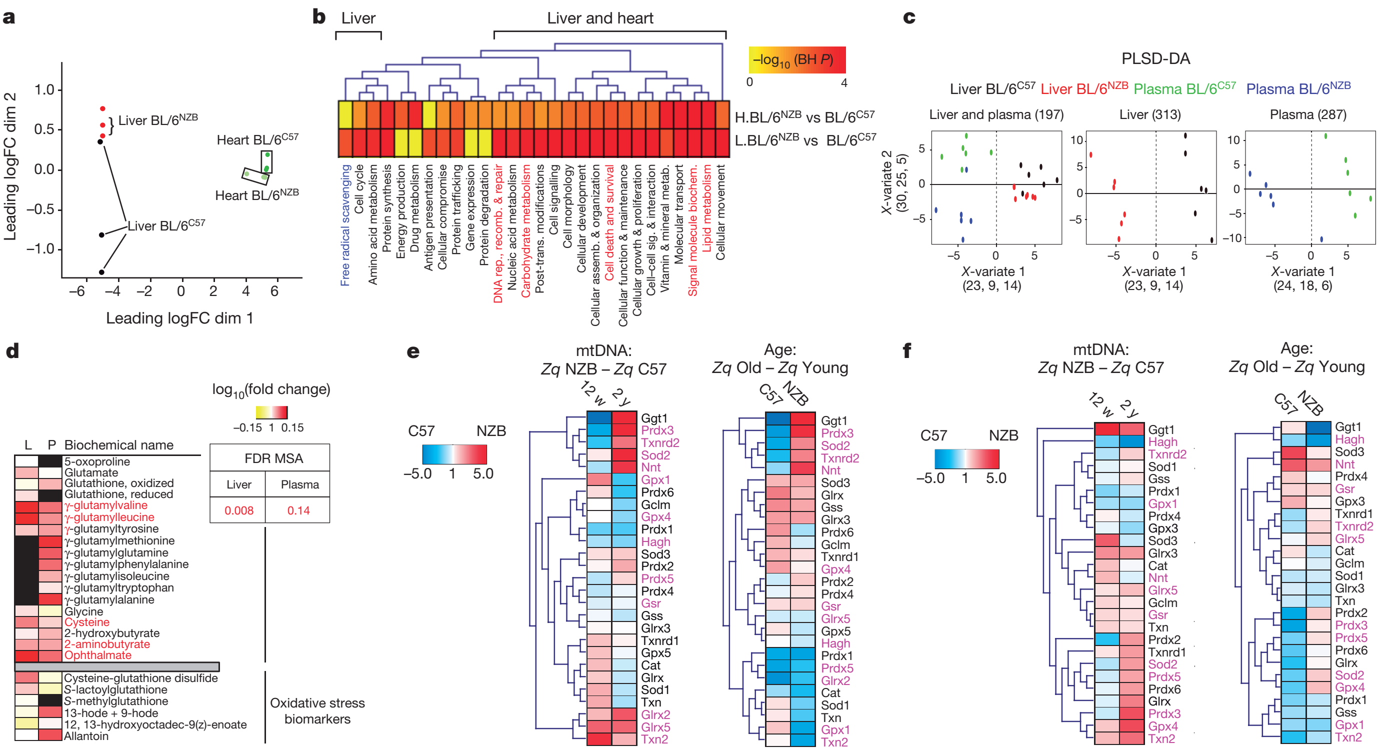


Figure 2 | Transcriptomic, metabolomics and proteomics adaptations to mtDNA haplotype. **a**, Multidimensional scaling (MDS) representation of RNA-seq analysis (10-weeks-old, $n = 3$ pools per genotype, 4 mice per pool). **b**, Heatmap and hierarchical clustering of enrichment ($-\log_{10}$ Benjamini–Hochberg (BH)-adjusted P values) of the functions defined by the ingenuity pathway analysis (IPA) in set of genes differentially expressed between conplastic mice in heart (H) (top) and liver (L) (bottom). Yellow, non-significant (BH $P = 1$); red, highly significant (BH $P \leq 10^{-4}$). **c**, Partial least squares regression discriminant analysis (PLS-DA) plot for plasma and liver metabolites (10-weeks-old, $n = 6$ per genotype). **d**, MSEA of all detected liver (L) and plasma (P) GSH metabolites regardless of relevance to PLS-DA (red text denotes PLS-DA relevant), and of selected oxidative stress metabolic biomarkers. **e**, **f**, Heat maps and hierarchical clustering of enrichment (Zq differences between indicated genotypes and/or ages) of the proteins quantified by iTRAQ involved in ROS metabolism in either liver (**e**) or heart (**f**). See Supplementary Data 4 for identity of different proteins and Zq values. Purple text denotes mitochondrial location ($n = 3$ per genotype and age).

cytosolic gamma-glutamyltranspeptidase 1 (involved in glutathione pool maintenance) and four mitochondrial redox regulators (Prdx3, Txnrd2, Sod2 and Nnt) were decreased in young BL/6^{NZB} mice but then strongly increased on ageing (Fig. 2e). Ageing has profound but different effects on the ROS handling proteomes of the two conplastic mouse lines (Fig. 2e).

Measurement of mitochondrial ROS *in vivo* using MitoB[13] revealed elevated ROS levels in young BL/6^{NZB} mice in both liver and heart (Fig. 3a). As shown in cultured cells[14], high levels of mitochondrial ROS correlated with mtDNA copy number (Fig. 3b). Chronic treatment of mice with *N*-acetyl cysteine (NAC) reduced the BL/6^{NZB} MitoP:MitoB ratio (Fig. 3a); equalized mtDNA copy number (Fig. 3b); and reduced respiration and ATP synthesis rates in isolated liver mitochondria (Fig. 3b). The higher rate of mitochondrial ROS in young BL/6^{NZB} animals (Fig. 3a) seems to conflict with their superior functional preservation with age. However, mitochondrial ROS production was inverted in animals above 30 weeks of age, with BL/6^{NZB} liver mitochondria from middle-aged mice producing similar levels of ROS to young animals, whereas BL/6^{C57} mitochondria showed a sharp, age-related increase in ROS (Fig. 3c).

There were no dramatic differences in mitochondrial complex or supercomplex assembly (Extended Data Fig. 6a), and both strains showed the expected loss of supercomplex CIII:CIV due to the mutant form of SCAF1 characteristic of the BL/6 nuclear background[15]. However, BL/6^{NZB} mice of all ages showed a slightly but consistently lower level of CI:CIII superassembly (Extended Data Fig. 6a, b).

Quantitative proteomics confirmed consistent behaviour patterns and between-genotype differences for structural proteins of all respiratory complexes, independently of whether they are encoded by nuclear or mtDNA genes (Extended Data Fig. 6c). In agreement with the differences in respiration (Fig. 3b), CI was less abundant in liver of young BL/6^{NZB} mice (Extended Data Fig. 6c). In parallel with the age-related decline in respiration in BL/6^{C57} mice (Fig. 1h), the amounts of respiratory complexes in liver decrease in BL/6^{C57} but not in BL/6^{NZB} mice (Extended Data Fig. 6c). An increase in the proportion of free CIII compared to that superassembled with CI is associated with adaptation to a fuel switch from glucose to fatty acids[16]. Consistent with this, the proteomic analysis revealed liver content changes in glycolytic and β-oxidization enzymes (Extended Data Fig. 6d).

Despite the influence of mtDNA haplotype on glucose and fatty acid metabolism, young mice showed no inter-strain differences in plasma glucose or in weight gain. In contrast, in one-year-old mice, the fasting-induced reduction in plasma glucose was significantly lower in BL/6^{C57} mice (Fig. 3d). This difference correlated with a lower capacity of BL/6^{C57} mice to regulate insulin levels (Fig. 3e) and to clear plasma glucose (Fig. 3f). Furthermore, on a high-fat diet (HFD), young BL/6^{C57} mice showed greater increases in body weight and white adipocyte size (Extended Data Fig. 7b, c). BL/6^{C57} mice showed an age-related increase in blood cholesterol, whereas levels in BL/6^{NZB} remained steady (Extended Data Fig. 7a)

The mtDNA haplotype influence in lipid and glucose metabolism was confirmed by *in vivo* magnetic resonance spectroscopy (MRS) analysis of lipid content in liver, heart and skeletal muscle. Conplastic animals showed no differences in lipid content at 20 weeks (Fig. 3g); however, cardiac fibres of BL/6^{NZB} mice had a significantly higher content and size of lipid droplets, suggesting a more

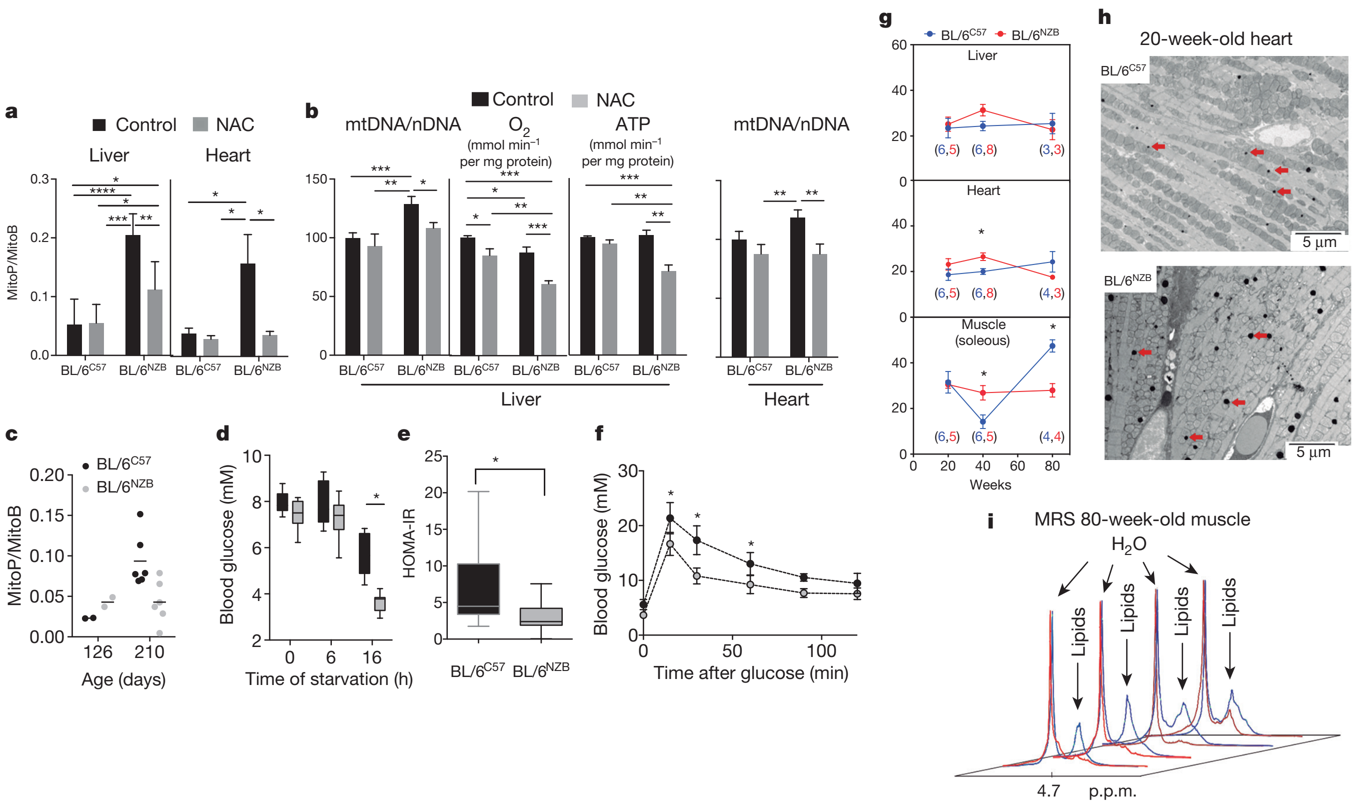


Figure 3 | Functional differences among conplastic strains. a, Relative amount of H_2O_2 ± NAC. **b**, Mitochondrial performance ± NAC. Liver: relative mtDNA amount; oxygen-dependent (coupled) respiration; pyruvate + malate-driven ATP production in isolated mitochondria. Heart: relative mtDNA amount. (12-week-old animals; $n = 6$ per genotype). **c**, Relative liver mitochondrial H_2O_2 generation (126 days old, $n = 2$ per genotype; 210 days old, $n = 6$ per genotype). **d**, Blood glucose after the indicated period of starvation (14 months old, $n = 7$ per genotype). **e**, The homeostasis model for insulin resistance (HOMA-IR) in overnight-starved mice (14 months old, $n = 12$ per genotype). **f**, Glucose tolerance test in starved mice (14 months old, $n = 7$ per genotype) **g**, Longitudinal quantification of lipid content by ^{1}H MRS for the indicated tissue, age and genotype (number of individuals in brackets). **h**, Representative heart TEM images illustrating differences in mitochondria and lipid droplets between 20-week-old conplastic mice ($n = 3$ per genotype). **i**, Quantification of lipid content by ^{1}H MRS in skeletal muscle (soleus) of 80-week-old animals ($n = 4$ per genotype). Data are means ± s.d., *$P < 0.05$, two-tailed *t*-test.

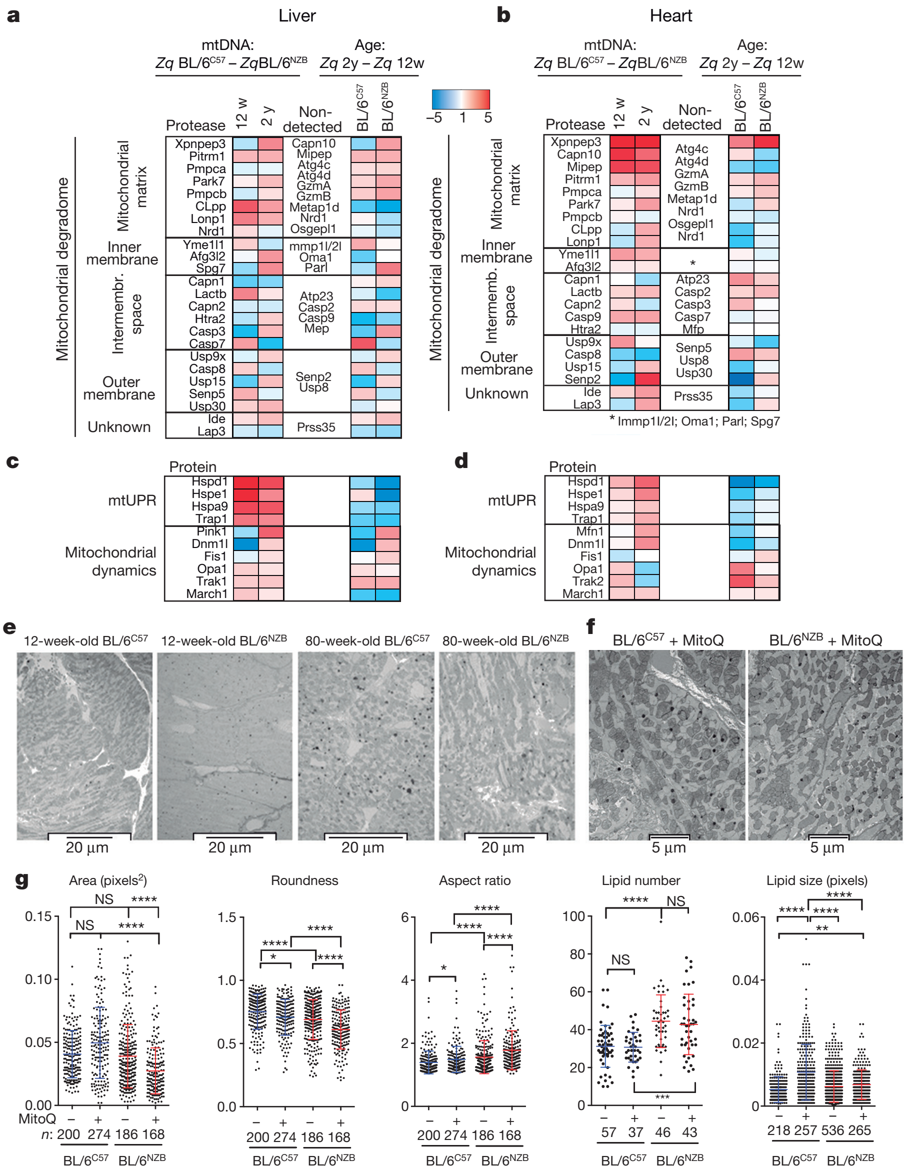


Figure 4 | Proteostasis, mtUPR and dynamics. See also Supplementary Data 4. **a**, **b**, Proteomic analysis of the mitodegradome. **c**, **d**, Detected proteases involved in the mitochondrial UPR and mitochondrial dynamics. Data are expressed as *Zq* differences between BL/6^{NZB} and BL/6^{C57} mice aged 12 weeks (w) or 2 years (y) (mtDNA) or between old and young animals of the same genotype (age). **e**, Representative TEM images of mouse left ventricle from 12-week-old (3 animals per genotype) and 80-week-old mice (2 animals per genotype). **f**, Representative TEM of cardiomyocytes from 12-week-old animals treated with MitoQ ($n = 3$ per genotype). **g**, Morphometric analysis of mitochondria in TEM images of left ventricular cardiomyocytes obtained from 3 animals per group. Data are given as means ± s.e.m., non-parametric Kruskal–Walls test (not normal distribution), $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.00001$; NS, not significant.

active mobilization of lipids to the heart (Fig. 3h). The MRS analysis revealed that mature BL/6^{NZB} adults (40 weeks) had higher tissue lipid content despite their lower body weight (Fig. 3g). At 80 weeks, lipid content in both strains was similar in liver and heart but was significantly higher in BL/6^{C57} muscle (Fig. 3g, i). This may reflect the accelerated deterioration of older BL/6^{C57} mice, with replacement of muscle mass by fat.

The mitochondrial degradome (mtDg) is involved in protein synthesis regulation, quality control, biogenesis, dynamics, mitophagy, and apoptosis[17]. Quantitative proteomics revealed between-strain differences in the mitochondrial degradome, especially in the heart, that were strongly influenced by age (Fig. 4a, b). Matrix proteases and chaperones involved in the mitochondrial unfolding response (mtUPR)[18] were elevated in BL/6^{NZB} animals, suggesting more active proteostasis (Fig. 4c, d). This was especially prominent in the liver, where elevation in stress chaperones was accompanied by elevation of proteases characteristic of the matrix (LonP1 and Clpp) and implicated in mtUPR (Fig. 4a, c).

The conplastic strains showed differences in critical components of the apparatus involved in mitochondrial dynamics (Fig. 4c, d). Heart transmission electron microscopy (TEM) revealed between-strain differences in mitochondrial aspect and dynamics, affecting density, morphology and cristae area (Fig. 4e–g, Extended Data Fig. 8); with more mitochondrial fragmentation in hearts of old BL/6^{C57} mice (Fig. 4e). When young mice were treated with the mitochondria-specific antioxidant MitoQ[19], heart mitochondria were less rounded in BL/6^{NZB} mice (Fig. 4f) and had a decreased mitochondrial area (Fig. 4g). MitoQ did not affect lipid droplet-accumulation between BL/6^{C57} and BL/6^{NZB} cardiomyocytes, but increased droplet size in BL/6^{C57} animals (Fig. 4g).

By highlighting the influence of mtDNA and nuclear (nDNA) interplay, our results demonstrate that substitution with a different wild-type

mtDNA variant is sufficient to promote differences in mitochondrial function and a cellular adaptive response. This phenomenon is orchestrated by the complex network of mitochondrial stress response pathways in which proteostasis, mtUPR and ROS signalling lead to adaptive responses that impact on the organism's metabolic performance and ageing. The potential for non-pathological mtDNA variants to shape organismal metabolism has implications for the emerging new field of mitochondrial replacement therapy in human oocytes[20,21]. Thus, consideration should be given to the use of donor mtDNA that is a close genetic match to the original mtDNA of the acceptor oocyte[22].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 June 2015; accepted 2 June 2016.
Published online 6 July 2016.

1. Wallace, D. C. Mitochondrial DNA variation in human radiation and disease. *Cell* **163,** 33–38 (2015).
2. Ruiz-Pesini, E. *et al.* Human mtDNA haplogroups associated with high or reduced spermatozoa motility. *Am. J. Hum. Genet.* **67,** 682–696 (2000).
3. Brown, M. D. *et al.* The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup J. *Hum. Genet.* **110,** 130–138 (2002).
4. Baudouin, S. V. *et al.* Mitochondrial DNA and survival after sepsis: a prospective study. *Lancet* **366,** 2118–2121 (2005).
5. Amo, T. & Brand, M. D. Were inefficient mitochondrial haplogroups selected during migrations of modern humans? A test using modular kinetic analysis of coupling in mitochondria from cybrid cell lines. *Biochem. J.* **404,** 345–351 (2007).
6. Amo, T., Yadava, N., Oh, R., Nicholls, D. G. & Brand, M. D. Experimental assessment of bioenergetic differences caused by the common European mitochondrial DNA haplogroups H and T. *Gene* **411,** 69–76 (2008).
7. Roubertoux, P. L. *et al.* Mitochondrial DNA modifies cognition in interaction with the nuclear genome and age in mice. *Nature Genet.* **35,** 65–69 (2003).
8. Yu, X. *et al.* Dissecting the effects of mtDNA variations on complex traits using mouse conplastic strains. *Genome Res.* **19,** 159–165 (2009).
9. Sharpley, M. S. *et al.* Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. *Cell* **151,** 333–343 (2012).
10. Ross, J. M. *et al.* Germline mitochondrial DNA mutations aggravate ageing and can impair brain development. *Nature* **501,** 412–415 (2013).
11. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153,** 1194–1217 (2013).
12. Soga, T. *et al.* Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J. Biol. Chem.* **281,** 16768–16776 (2006).
13. Cochemé, H. M. *et al.* Using the mitochondria-targeted ratiometric mass spectrometry probe MitoB to measure H_2O_2 in living *Drosophila*. *Nature Protocols* **7,** 946–958 (2012).
14. Moreno-Loshuertos, R. *et al.* Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants. *Nature Genet.* **38,** 1261–1268 (2006).
15. Lapuente-Brun, E. *et al.* Supercomplex assembly determines electron flux in the mitochondrial electron transport chain. *Science* **340,** 1567–1570 (2013).
16. Guarás, A. *et al.* The $CoQH_2$/CoQ ratio serves as a sensor of respiratory chain efficiency. *Cell Rep.* **15,** 197–209 (2016).
17. Quirós, P. M., Langer, T. & López-Otín, C. New roles for mitochondrial proteases in health, ageing and disease. *Nature Rev. Mol. Cell Biol.* **16,** 345–359 (2015).
18. Houtkooper, R. H. *et al.* Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* **497,** 451–457 (2013).
19. Kelso, G. F. *et al.* Selective targeting of a redox-active ubiquinone to mitochondria within cells: antioxidant and antiapoptotic properties. *J. Biol. Chem.* **276,** 4588–4596 (2001).
20. Tachibana, M. *et al.* Mitochondrial gene replacement in primate offspring and embryonic stem cells. *Nature* **461,** 367–372 (2009).
21. Craven, L. *et al.* Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease. *Nature* **465,** 82–85 (2010).
22. Chinnery, P. F. *et al.* The challenges of mitochondrial replacement. *PLoS Genet.* **10,** e1004315 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. López-Otín for critical reading; C. Jimenez for management; the Transgenesis Unit of the CNIC and M. Cueva for mouse work, A. Molina and R. Doohan for histology, the Genomics Unit of the CNIC for RNA-seq; ICTS-CNME (UCM) for technical support; J. Amigo and X. Bello for NGS bioinformatics support; and S. Bartlett for English editing. Supported by grants from the MEyC (SAF2015-65633-R, CSD2007-00020, SAF2012-38449, BIO2012-37926 and BIO2015-67580-P), the Madrid (CAM/P2010/BMD-2402), the EU (PCIG10-GA-2011-304217 and FP7-PEOPLE-2012-ITN-GA-317433), the ISC III (PI09-00946, PI12/01297, PI13-01136 and PI11-00078, PRB2 (IPT13/0001, ProteoRed), RIC ((RD12/0042/0045) and RETICS (RD12/0042/00056)). The CNIC is supported by the MEyC and the Pro-CNIC Foundation and is a SO-Center of Excellence (SEV-2015-0505).

Author Contributions A.L.-P., R.M.-L., A.V.L.-V., R.A.-P. and S.C. performed experimental work. F.S.-C. and C.T. analysed genomic and transcriptomic data. E.A. and I.F. performed the telomere analysis. A.G.-G., M.L.B.-M. and E.R. performed the mouse phenotyping. A.L. and M.P.M. did the assessment of ROS levels. R.C., B.S. and A.C. analysed the microarrays and mtDNA next-generation sequencing. A.V.L.-V. and J.R.-C. analysed the metabolomics. E.C. and J.V. performed the proteomics. J.A.E., A.L.-P., R.M.-L., A.P.-M. and P.F.-S. contributed to design and analysis. A.L.-P. and J.A.E. wrote the manuscript. J.A.E. directed and designed the research.

Author Information RNA-seq data was deposited at GEO under accession number GSE56933. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.E. (jaenriquez@cnic.es).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Generation of conplastic mice and genotyping methodologies. Parental C57BL/6JOlaHsd and NZB/OlaHsd strains were purchased from Harlan Laboratories. The C57BL/6JOlaHsd strain was derived in 1974 from the C57BL/6J strain from Jackson Laboratories but differs from the parental strain. The C57BL/6JOlaHsd strain does not harbour the nicotinamide nucleotide transhydrogenase (NNT) spontaneous mutation that renders the encoded enzyme undetectable, and which is characteristic of the C57BL/6J strain provided by Jackson Laboratories. To avoid confusion, we call C57BL/6JOlaHsd mice C57BL/6 throughout the manuscript. Conplastic strains were obtained by backcrossing females (mitochondrial donors) with males of the parental recipient strain over 20 generations to create a new line harbouring the nuclear genome of one strain and the mtDNA of the other. Conplastic strains were maintained by backcrossing females (mitochondrial donors) of the conplastic strain with males of the parental recipient strain. Offspring after the tenth generation of backcrossing were considered conplastic. Total DNA extracted from tissue homogenates was genotyped by PCR restriction-fragment length polymorphism analysis (RFLP) and by Affymetrix Mouse Diversity Genotyping Array, which interrogates 623,000 single nucleotide polymorphisms (SNPs) covering the entire mouse genome at a resolution of 4.3 kb.

mtDNA next-generation sequencing. Liver samples were taken from 12- ($n = 5$ per genotype), 40- ($n = 4$ per genotype) and 80-week-old ($n = 5$ per genotype) conplastic mice. Total DNA was extracted with the DNeasy Blood and Tissue kit (Qiagen). mtDNA was amplified in two independent reactions using primers 7309Fw (CATAGGGCACCAATGATACTG) — 16265Rv (GAGTTTTGGTTCACGGAACAT) and 15358Fw (AAGAA GAAGGAGCTACTCCCC) — 8257Rv (CAGCTCATAGTGGAATGGCTA). Long-range PCRs were carried out using Takara Taq DNA Polymerase (Clontech). Products were cleaned using the QiAquick PCR Purification kit (Qiagen), quantified with the Qubit dsDNA BR Assay kit (Invitrogen) and mixed in equimolar amounts. mtDNA was then enzymatically sheared using the Ion Shear Plus Reagent (Ion Torrent) following the manufacturer's instructions. To yield fragments with a median size of 200 bp, the shearing reaction was adjusted to 8 min at 37 °C, and fragments were size selected using Agencourt AMPure XP Reagent (Beckman Coulter). Both ends of each fragment were ligated with Ion P1 Adapters and the Ion Xpress Barcode Adapters. The ligated DNA was purified using Agencourt AMPure XP Reagent. Quantification and qualification for the amplified libraries were quantified and their quality assessed with Agilent 2200 TapeStation System and the Agilent High Sensitivity D1000 Screen Tape (Agilent Technologies) and with the Ion Library TaqMan Quantitation Kit (Ion Torrent). All barcoded libraries were pooled in equimolar amounts (50 pM) to ensure equal representation of each barcoded library in the sequencing run. Template preparation and Ion PI Chip loading were performed on the Ion Chef Instrument. NGS was performed on Ion Proton platform (Ion Torrent) with Ion PI Chip.

All raw reads were aligned to the C57BL/6J mouse mtDNA reference sequence (AY172335.1) using Torrent Mapping Alignment Program (TMAP) version 5.0.7. Variants were called using the Torrent Variant Caller (TVC) version 5.0.7, applying a minimum allele frequency threshold of 0.001 and reducing the coverage to 10,000 in over-sampled locations (keeping default values for somatic mutations for the rest of parameters) and then filed into variant call format (VCF) reports. Additionally, the relative frequency of the alternative allele was calculated for each individual and position. Cases deviating more than five standard deviations from the average frequency in that position for the 28 individuals were compared with variants obtained from TVC. In this way false positives (positions identified as variants in TVC but not clearly deviating from the mean value) were identified and discarded from the final variant set.

At the same time, variants found in the complete set of NZB individuals but lacking in all C57 individuals were identified as sequence differences between NZB mtDNA and C57 mtDNA. Extended Data Fig. 1b was generated with Geneious version 9.1.2, created by Biomatters.

Mouse experimentation. All animal procedures conformed to EU Directive 86/609/EEC and Recommendation 2007/526/EC regarding the protection of animals used for experimental and other scientific purposes, enforced in Spanish law under Real Decreto 1201/2005. Approval the different experimental protocols require the estimation of the adequate sample size as well as the definition of the randomization and blinding criteria. Mice under endpoint criteria were excluded from analysis in long-term experiments. The mice were fed a standard chow diet (5K67 LabDiet) or, when indicated, a high-fat diet (D12492 Open Source Diets). Unless specified, experiments were carried out with males.

Growth curves. To calculate growth curves, mice were weighed at the starting age and every week/month thereafter for the period indicated in each experiment. Three independent groups of animals were used for Fig. 1b and Extended Data Figs 2a and 6b.

Survival study. Mice found dead at each daily inspection were considered as censored deaths, and were necropsied to tumour inspection. Criteria for euthanasia were based on an independent assessment made by a veterinarian, and only cases where the condition of the animal was considered incompatible with continued survival are represented as deaths in the curves. Survival curves, weight gain and tumour incidence analysis (Fig. 1a, b, f) were done in the same group of animals.

Behaviour and metabolism analysis. To analyse food and water intake, an independent group of 10–12 males per strain were individually housed, and weekly intake was calculated as the difference between the supplied weight of food and drink and the weight remaining after one week. Intake was calculated in males because females were grouped and had to compete for food and drink. Mouse activity was evaluated in 10-week-old mice using the PHYSIOCAGE system (Panlab). Animals were acclimatized to the room and cages in the week before the experiment, and measurements were made on four consecutive days (five days and four nights). Data were analysed with Metabolism v2.2 (Panlab). The first diurnal phase (12 h) was considered part of the acclimatization period and was eliminated from the final result. Indirect calorimetry measurements were made of individual 10-week-old mice (10 per sex and strain) housed in Oxylet System metabolic chambers (Panlab Harvard Apparatus). Before experiments, mice were acclimatized in the cages for 1 week. Throughout the 24 h measurement period, oxygen consumption and carbon dioxide production rates were measured at 24-min intervals, with a 4 min sample period and a 1 min purge; simultaneous measurements were made in four metabolic chambers. Respiratory exchange rate (RER, or respiratory quotient, RQ), a measure of metabolic substrate choice, was calculated as the ratio between CO_2 production and O_2 consumption. Energy expenditure (EE) was calculated as $(3.815 + (1.232 \times RQ)) \times VO_2 \times 1.44)$. Data were analysed with Metabolism v2.2 (Panlab). Investigators were not blinded to the group allocation during the experiment and the sub-groups of animals included in repeated studies was randomly organized.

Glucose metabolism and blood biochemistry analysis. Experiments were performed in an independent group of 12 mice per genotype. Mice fed the chow diet were fasted for 6 or 16 h with free access to water. For intraperitoneal (IP) glucose tolerance tests (GTT), the mice received glucose injections of 1 or 2 mg per kg. Mice were bled from a tail clip. Blood glucose was measured with a handheld glucometer before injection (time 0) and at 15, 30, 60, 90 and 120 min after injection. Insulin was measured by an enzyme-linked immunosorbent assay (ELISA) (Millipore), 3 technical replicates per sample. The homeostasis model for insulin resistance (HOMA-IR) was calculated from the fasting blood glucose (mmol per litre) × fasting plasma insulin ($\mu U\ ml^{-1}$) divided by 22.5.

Levels of serum cholesterol (ChoT), low-density lipoprotein (LDL), and high-density lipoprotein (HDL) were determined by biochemical analysis of submandibular blood samples. Serum biochemistry was conducted in a Dimension RxL Max automated analyser. Investigators were not blinded to the group allocation when assessing the outcome.

Histological analysis. Tissue samples were fixed in 4% paraformaldehyde (24 h), processed and embedded in paraffin. Sections (5 μm) were prepared and mounted on coverslips for staining with haematoxylin and eosin and tissue images were captured using a NanoZoomer 2.0RS digital slide scanner (Hamamatsu). Sections of epididymal fat pad stained for white adipose tissue were examined under a Nikon 90i microscope fitted with a 10× objective. Adipocyte cross-sectional area was determined from epididymal fat pads using ImageJ software (4 sections per animal). Slides from liver necropsies were digitalized and analysed with NDP.view2 viewing (Hamamatsu). Lipofuscin area was determined from digitalized haematoxylin and eosin histological ovary sections (3 sections per ovary) using NDP.view2. Measurements were obtained by applying colour thresholds with ImageJ 1.48v (NIH). Investigators were blinded to the group allocation when assessing the outcome.

***In vivo* ROS determination.** *In vivo* mitochondrial hydrogen peroxide was measured using the MitoB mass spectrometric probe as described previously[13,23]. Briefly, 25 nmol MitoB was injected into the tail vein between 8 a.m. and 10 a.m. After 6 h, mice were euthanized by cervical dislocation and tissues were harvested and flash frozen in liquid nitrogen. Tissues were stored at −80 °C and transported to Cambridge, UK on dry ice for analysis. When indicated, mice were treated with 1% NAC in drinking water for 8 weeks before MitoB injection. Investigators were blinded to the group allocation when assessing the outcome.

OXPHOS function. Mitochondria were isolated from mouse liver samples as described[15], and oxygen consumption and ATP production were measured[15]. For mtDNA copy number quantification, total DNA was isolated from tissue homogenate preparation (liver and heart) and was quantified by real-time PCR using 7900HT Fast Real-Time PCR System (Applied Biosystem) as described[14]. When indicated, mice were treated with 1% NAC in drinking water for 8 weeks before

sample collection. Each data represent mean of 4 technical replicates. Investigators were not blinded to the group allocation when assessing the outcome.

Blue native electrophoresis (BNE). Supercomplex levels and composition were analysed in isolated mitochondria from liver and heart by BNE as described[24] using the following antibodies: core1, ndufa9 (Abcam); FpSDH, COXI (Novex) and Tom20 (Santa Cruz Biotech). Investigators were not blinded to the group allocation when assessing the outcome.

Telomere length analysis. Telomere length was measured in tail skin samples by quantitative fluorescence *in situ* hybridization (Q-FISH). Paraffin-embedded tissue sections (7 μm) were hybridized with a Cy3-labelled PNA probe containing the telomere repeat sequence[25]. DAPI and Cy3 signals were acquired in separate channels using a confocal microscope (Leica SP5) fitted with a 60× objective. All slides were stained at the same time, and images were acquired consecutively, with lasers kept at the same intensity to avoid day-to-day variations. Maximum intensity projections from 12-bit image stacks (15 sections at 1.0 μm steps) were generated for image quantification.

The DAPI images were signal-intensity thresholded and segmented using ImageJ. Quantitative image analysis was performed using the Metamorph platform (Molecular Devices)[25]. After conversion to a 1-bit binary mask, the DAPI image was used to define the nuclear area. The binary DAPI mask was applied to the Cy3 image to obtain a combined image with the telomere fluorescence information for each nucleus. Cy3 fluorescence intensity was measured in arbitrary fluorescence units (au) as the mean grey value in each nucleus. Only cells from the bulge of the hair follicle were analysed. Telomere intensity values were exported to Excel. The analysis was conducted on one-year-old BL/6^{C57} ($n = 6$), one-year-old BL/6^{NZB} ($n = 6$), two-year-old BL/6^{C57}($n = 4$), and two-year-old BL/6^{NZB} ($n = 5$) mice, and two images were acquired per animal. The number of cells analysed per group is indicated in the frequency histograms (Fig. 1i).

The statistical significance of differences in telomere length distributions was calculated with a Wilcoxon's rank sum test. A two-tailed *t*-test was used to compare the reduction in mean telomere length and the percentages of short and long telomeres. Calculations were made in Microsoft Excel and Prism. Differences were considered significant at $^{*}P < 0.05$, at $^{**}P < 0.01$, and at $^{***}P < 0.001$. Investigators were not blinded to the group allocation when assessing the outcome.

Phenotype scoring. Mice aged ~2 years old were scored for a variety of ageing markers (reduced hair density, greying hair, kyphosis, body weight reduction and development of systemic neoplasias) as previously described[10]. Investigators were blinded to the group allocation during the experiment.

Transcriptomic analysis. Twelve 10-week-old males per genotype were killed by cervical decapitation, and liver (median lobe) and heart (left ventricle) tissues were dissected and immediately frozen in liquid nitrogen. Total RNA was extracted with TRIzol reagent and then purified on RNeasy spin columns (Qiagen). RNA integrity (RNA Integrity Score ≥ 7.9) and quantity were determined with an Agilent 2100 Bioanalyzer. Equal RNA amounts from individual animals were combined at 4 samples per pool, and 3 pools per genotype were analysed in the Illumina Genome Analyser IIx. Reads were pre-processed with Cutadapt 1.2.1 to remove the TruSeq adaptor. To focus on nuclear genes, we identified and removed mitochondrial reads by aligning all reads to mitochondrial DNA using Tophat. Resulting reads were mapped on the mouse transcriptome (Ensembl gene-build GRCm38.v70) and genome, using RSEM v1.2.3 and Bowtie2 v2.0.6. Genes with at least one count per million in at least three samples were considered for downstream analysis. As data were generated in two different sequencing runs, we used the R function Combat (http://www.bu.edu/jlab/wp-assets/ComBat/Download.html) to adjust for potential batch effects. Statistical analysis was performed on the remaining set of genes using the bioconductor package EdgeR[26]. Raw reads and TMM-normalized batch corrected counts have been deposited in GEO with the accession number GSE56933. Sets of genes differentially expressed across conditions were analysed for functional associations using IPA (Ingenuity Knowledge Database, http://www.ingenuity.com). Benjamini–Hochberg adjusted *P* values are reported for differential gene expression and IPA functional analysis. Gene Set Enrichment Analysis (GSEA[27]) identified sets of co-expressed genes that are functionally linked based on different ontologies and/or pathways. Gene sets with a FDR < 0.25 will be considered to functionally related. Investigators were not blinded to the group allocation when assessing the outcome.

Metabolomic analysis. Samples of liver and plasma from 10-week-old mice ($n = 6$ per genotype) were collected, frozen in liquid nitrogen and sent to Metabolon (Durham, NC, USA) for metabolomics analysis according to their proprietary methodology. The analysis profiled 313 metabolites in liver samples and 287 in plasma, with 197 common to both sources.

For the multivariate model analysis, between-group differences in the abundance of metabolites identified by mass spectrometry (LC–MS and GC–MS) were assessed by Welch's two-sample *t*-test. Differences were considered statistically significant at $P < 0.05$. Chemometric statistical analysis was performed using R scripts (http://cran.r-project.org/) and the mixOmics package (v. 5.0.1) executed in RStudio (v. 0.98.501) for MacosX. Most metabolomic analysis employs unsupervised principal components analysis (PCA) to explore the quality of data and detect possible outliers and a supervised partial least squares-discriminant analysis (PLS-DA) to determine the number of metabolites responsible. PLS-DA is a supervised extension of PCA used to distinguish two or more classes by searching for original metabolic variables (*X* matrix) that are correlated to class membership (that is, liver, plasma, and mouse model) (*Y* matrix). Results of the PCA and PLS-DA were visualized by a scatter score plot of different PCs (PCA) and the first two latent variables (PLS-DA). The original *X* matrix contains the metabolic variables, while the *Y* matrix contains the class variable for which values are chosen to be the class descriptor. We performed the analysis by comparing BL/6^{C57} samples with BL/6^{NZB} samples, and liver samples with plasma samples. False discovery rate and overfitting were avoided by using a leave-one-out cross-validation, performed with the calibration data set, to determine the number of latent variables for each fold and to construct a model. The resulting model was applied to the test data set, including the 287 plasma metabolites and the 313 liver metabolites, to compute the predicted Y values. Due to the small number of samples, this process was repeated the number of times needed in each analysis so that every sample served as an independent test sample once and only once. The predicted Y values produced by the cross-model validation were used to examine the classification error and determine the number of most discriminant metabolites. Based on this, we selected only those metabolites that appear in at least 50% of the repeated analysis. For metabolomic set enrichment analysis (MSEA)[27], raw data from the metabolic analysis were normalized by quantiles using the Bioconductor LIMMA package. Metabolites with missing values were excluded. A moderated *t*-test, also implemented in the LIMMA package, was used to calculate log fold-differences between male BL/6^{NZB} and BL/6^{C57} mice for liver and plasma samples separately, together with raw and Benjamini–Hochberg adjusted *P* values. MSEA was then run on the normalized samples and the signal-to-noise ratio statistic was used to rank metabolites in each pathway of interest. A false discovery rate (FDR) < 0.25 was considered significant. We considered sets of at least 10 metabolites in all analyses, except for bile acid, which required at least 5 metabolites to be in the same class for a result to be considered significant.

Proteomic analysis. Heart and liver protein extracts from 12- or 80-week-old BL/6^{C57} and BL/6^{NZB} mice ($n = 3$ per genotype and age) were reducted, alkylated with iodoacetamide and digested with trypsin and the resulting peptides were labelled with iTRAQ8-plex, according to the manufacturer's instructions. Labelled peptides were injected onto a C-18 reversed phase (RP) nano-column (75 μm I.D. and 50 cm, Acclaim PepMap, Thermo Fisher, San José, CA, USA) and analysed in a continuous acetonitrile gradient consisting of 8–31% B for 240 min, and 50–90% B for 1 min (B = 0.5% formic acid in acetonitrile). Peptides were eluted from the RP nano-column at a flow rate of ~200 nl min^{-1} to an emitter nanospray needle for real-time ionization and peptide fragmentation in a Q-Exactive HF mass spectrometer (Thermo Fisher). Mass spectra were acquired in a data-dependent manner, with an automatic switch between MS and MS/MS using a top 20 method.

The raw files were analysed with Proteome Discoverer (version 1.4, Thermo Fisher Scientific), using a Uniprot database containing a joint Mouse + Rat Swissprot database (Uniprot release 57.3 May 2009; 26.885 entries). For database searching, parameters were selected as follows: trypsin digestion with 2 maximum missed cleavage sites, precursor mass tolerance of 800 p.p.m., fragment mass tolerance of 30 mmu, variable methionine oxidation and fixed cysteine carbamidomethylation, lysine and N-terminal modification of + 304.2054 (iTRAQ label). The same collections of MS/MS spectra were also searched against inverted databases constructed from the same target databases. Peptide identification from MS/MS data was performed using the probability ratio method[28]. FDR of peptide identifications were calculated using the refined method[29]; 1% FDR was used as the criterion for peptide identification. Each peptide was assigned only to the best protein proposed by the Proteome Discoverer algorithm. Quantitative information was extracted from iTRAQ sample MS/MS spectra using an in-house developed program (QuiXoT) and protein abundance changes were analysed using the Generic Integration Algorithm (GIA), as described in ref. 30. The GIA was also used to integrate the quantitative protein information from several experiments into a unique protein value. Calculation of statistical weights of each quantitation at the spectrum level was performed according to the WSPP model. The validity of the null hypothesis at each one of the levels (spectrum, peptide, protein within an experiment and protein) was carefully checked by plotting the cumulative distributions. Three mice for each genetic background and age were used as biological replicates for heart experiments, and two mice were used for liver experiments. Changes in relative protein abundance are expressed using the standardized $\log_2$ ratio (*Zq*) variable[30]. Investigators were not blinded to the group allocation when assessing the outcome.

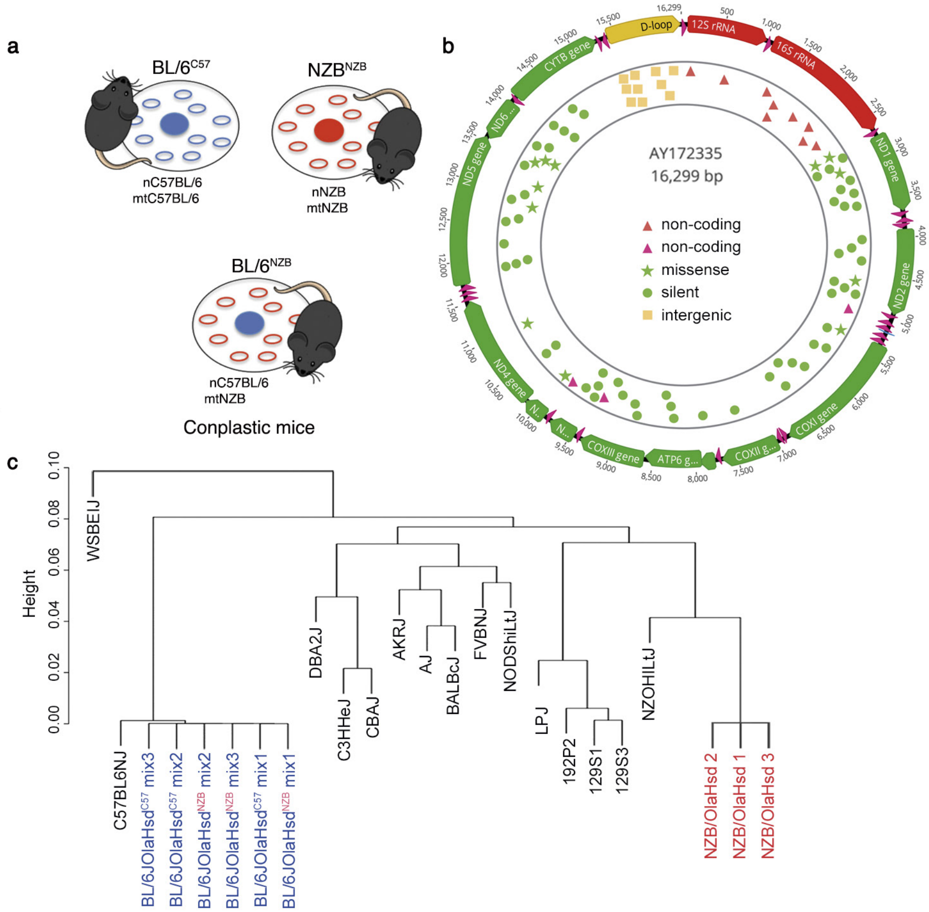
^{1}H magnetic resonance spectroscopy (MRS). Male BL/6^{NZB} and BL/6^{C57} mice (80 weeks old ($n = 4$–5), 20 weeks old ($n = 6$–8)) weighing 30–50g each were housed in groups of 4 with unrestricted access to food and water. Spectroscopy examinations were performed with a 7T preclinical system (Agilent Varian, Palo Alto, USA) equipped with a DD2 console and an active shielded 205/120 gradient insert coil with a 130 mT per m maximum strength. A double-tuned circular transmit/receive coil was used for phosphorus/proton (20 mm), located mainly on the soleus muscle, heart and liver (Rapid Biomedical GmBH, Rimpar Germany). Investigators were blinded to the group allocation during the experiment.

Proton spectra were acquired by 128 transients with 2,048 complex points with a spectral bandwidth of 10 kHz and a repetition time of 1.2 ms. Spectra were acquired with adiabatic radiofrequency pulses to improve sensitivity and minimize spectral distortions with an Ernst flip angle. Proton chemical shifts were expressed relative to the water signal (4.7 p.p.m.). Signals in nuclear magnetic resonance spectra were determined quantitatively on a Macintosh computer by integration after automatic or manual baseline correction, with fitting spectrum peak (after phase and baseline correction) to a Lorentzian function using the Mestrenova program (Mestrelab Research, Santiago de Compostela, Spain; released 2015-02-04 version:10.0.1-14719). An exponential line broadening (3 Hz for proton) was applied before Fourier transformation.

Transmission electron microscopy (TEM). Mice were euthanized by cervical dislocation and the left ventricle apex was immediately dissected (12 weeks old, $n = 3$ per genotype and 80 weeks old, $n = 2$ per genotype). Heart samples were fixed in 2.5% glutaraldehyde, 4% formaldehyde in 0.1 M HEPES buffer for 4–5 h. After buffer washes, samples were post-fixed for 1 h at room temperature in a 1:1 solution of 1% osmium tetroxide and 3% aqueous potassium ferrocyanide. Samples were rinsed in distilled H_2O. Tissues were dehydrated through a graded acetone series and embedded in Spurr's low viscosity embedding mixture (Electron Microscopy Sciences). Ultra-thin sections (60 nm) were then mounted on copper grids and stained with lead citrate. Samples were examined on a JEOL 10-10 electron microscope through 1,500×, 5,000×, 40,000× and 80,000× objectives. Mitochondrial morphometry, cristae area (2 images, 5,000×, per animal) and lipid droplets (50 images, 5,000×, per animal) were segmented manually and analysed using Fiji (http://fiji.sc/Fiji) and ImageJ 1.48v software. The investigator was blinded to the group allocation when assessing the outcome.

Statistical analysis. Unless specified, statistical analyses and graphics were produced with GraphPad Prism 6 software. Data sets were compared by unpaired two-tailed t-tests. Differences were considered statistically significant at P values below 0.05. $*P < 0.05$; $**P < 0.005$; $***P < 0.0005$. All results are presented as mean ± s.d. or mean ± s.e.m.

23. Logan, A. *et al.* Using exomarkers to assess mitochondrial reactive species *in vivo*. *Biochim. Biophys. Acta* **1840,** 923–930 (2014).
24. Acín-Pérez, R., Fernández-Silva, P., Peleato, M. L., Pérez-Martos, A. & Enriquez, J. A. Respiratory active mitochondrial supercomplexes. *Mol. Cell* **32,** 529–539 (2008).
25. Flores, I. *et al.* The longest telomeres: a general signature of adult stem cell compartments. *Genes Dev.* **22,** 654–667 (2008).
26. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
27. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
28. Martínez-Bartolomé, S. *et al.* Properties of average score distributions of SEQUEST: the probability ratio method. *Mol. Cell. Proteomics* **7,** 1135–1145 (2008).
29. Navarro, P. & Vázquez, J. A refined method to calculate false discovery rates for peptide identification using decoy databases. *J. Proteome Res.* **8,** 1792–1796 (2009).
30. García-Marqués, F. *et al.* A novel systems-biology algorithm for the analysis of coordinated protein responses using quantitative proteomics. *Mol. Cell. Proteomics* **15,** 1740–1760 (2016).



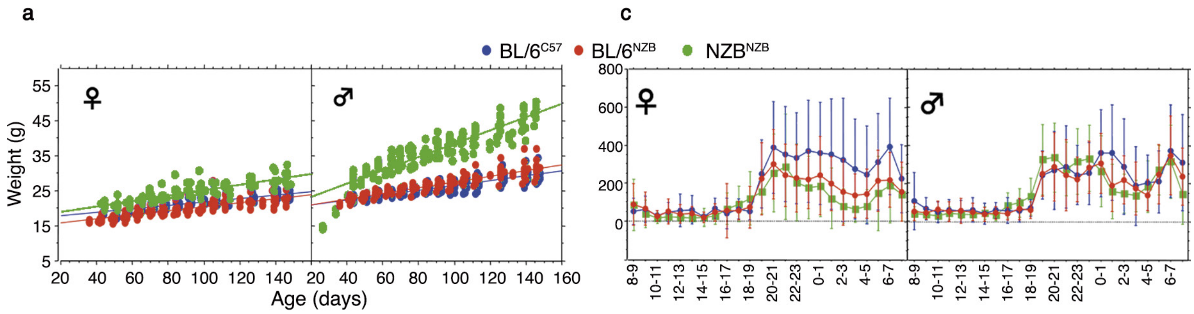
d

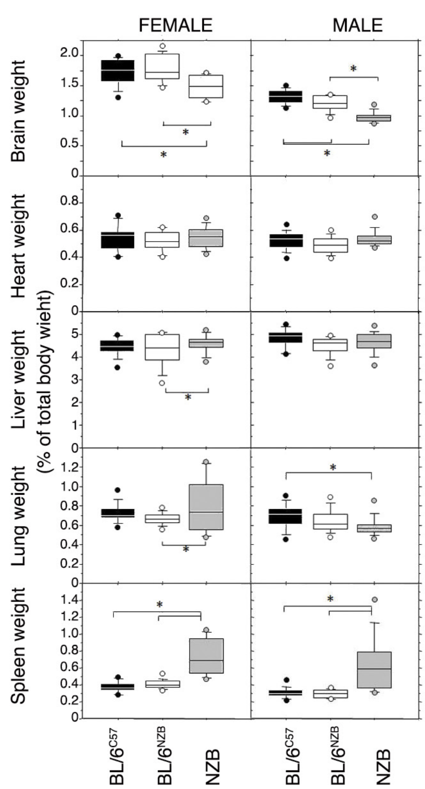
Comparison		Nº SNPs	Total
BL/6^{C57} vs. BL/6^{NZB}	Same	105,775	105,972
	Different	197 (0.1%)	
NZB^{NZB} vs. BL/6^{NZB}	Same	30	6,508
	Different	6,478 (99.5%)	

e

	BL/6^{C57} 1	BL/6^{C57} 2	BL/6^{C57} 3	BL/6^{NZB} 1	BL/6^{NZB} 2	BL/6^{NZB} 3	NZB^{NZB} 1	NZB^{NZB} 2
BL/6^{C57} 2	0.3 n=10243							
BL/6^{C57} 3	0.2 n=9280	0.2 n=8902						
BL/6^{NZB} 1	0.3 n=9921	0.4 n=907	0.4 n=8410					
BL/6^{NZB} 2	0.4 n=9755	0.3 n=9090	0.3 n=8356	0.3 n=8739				
BL/6^{NZB} 3	0.5 n=13326	0.5 n=11276	0,4 n=10197	0.4 n=10897	0.4 n=10674			
NZB^{NZB} 1	86.2 n=8666	84.7 n=7782	85.8 n=7109	85.7 n=7631	85.5 n=7466	83.3 n=9874		
NZB^{NZB} 2	86.2 n=6605	84.4 n=6340	85.1 n=5966	85.8 n=6137	85.5 n=6116	84.7 n=7214	0.7 n=6570	
NZB^{NZB} 3	86.7 n=6731	84.9 n=6462	85.6 n=6050	86.0 n=6323	85.5 n=6294	84.9 n=7429	0.7 n=6716	0.4 n=6716

Extended Data Figure 1 | Conplastic mouse generation and nuclear genome identity. a, Schematic of the conplastic animal concept. **b**, Sequence differences between NZB and C57 mtDNAs by next-generation sequencing. See also Extended Data Table 1. **c**, Cluster dendrogram based on the expressed genome of the indicated mouse strain and the RNA-seq-derived expressed genome of conplastic animals. **d**, Comparison of nuclear single nucleotide polymorphism (SNP) differences between the indicated mouse lines by Affimetrix Mouse Diversity Genotyping Array analysis. **e**, Percentage of genotype differences between the mouse nuclear DNA SNPs identified by RNA-seq; the number of SNPs compared is indicated.





d

Area under activity curves

Gender	Strain	Day Value	Day *p*	Night Value	Night *p*
Male	BL/6^{C57}	786.2	0.694	1795.0	0.107
	BL/6^{NZB}	726.7		1511.0	
	NZB^{NZB}	633.9		1440.0	
Female	BL/6^{C57}	621.9	0.692	2070.7	***0.001***
	BL/6^{NZB}	682.1		1535.3	
	NZB^{NZB}	512.2		973.0	

e

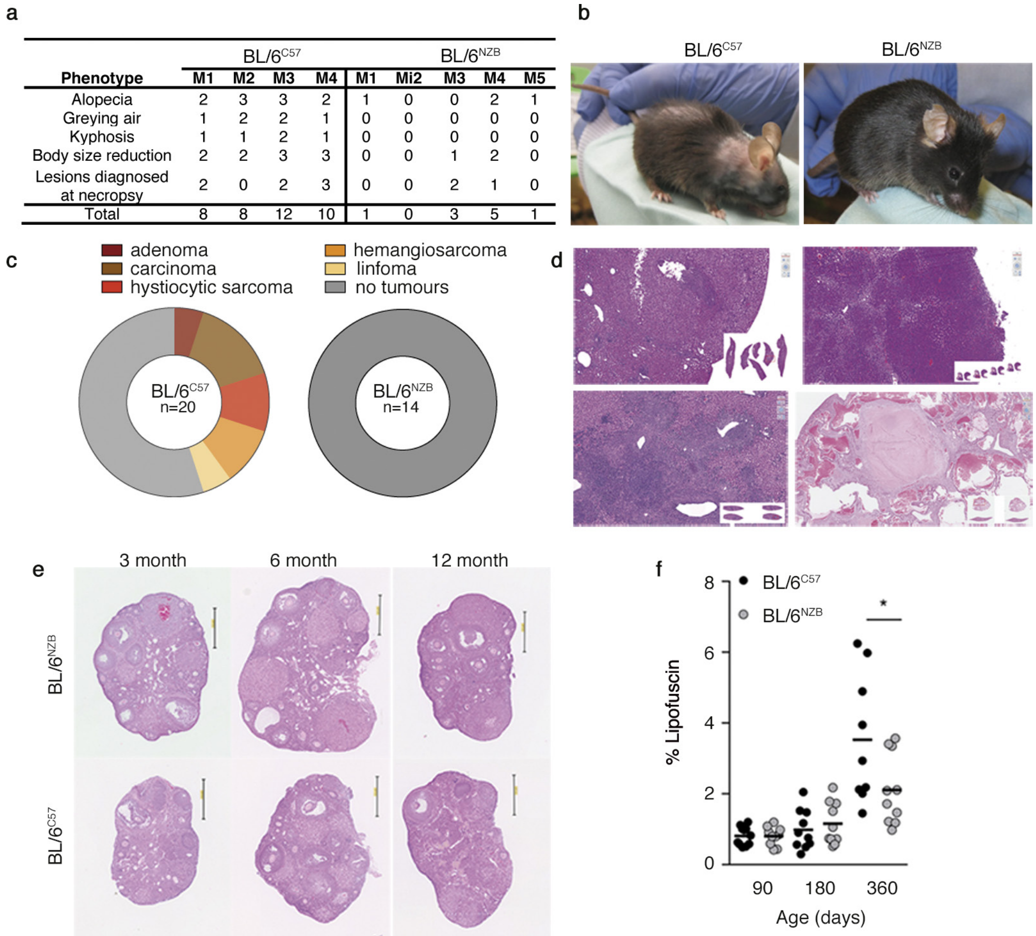
Activity vs. respiratory parameters

Gender	Strain	Activity/O_2cons.	Activity/CO_2prod.	Activity/EE
Male	BL/6^{C57}	211.50±90.40*(*)*	208.51±87.19*(*)*	29.03±12.32*(*)*
	BL/6^{NZB}	162.99±38.88	166.43±46.13	22.58±5.56
	NZB^{NZB}	134.91±51.00	139.04±59.44	18.65±7.25
Female	BL/6^{C57}	247.49±104.00*(#)*	259.69±107.61*(#)*	34.60±14.42*(#)*
	BL/6^{NZB}	156.51±78.64	164.70±82.20	21.83±10.94
	NZB^{NZB}	119.32±87.36	135.59±107.29	16.85±12.42

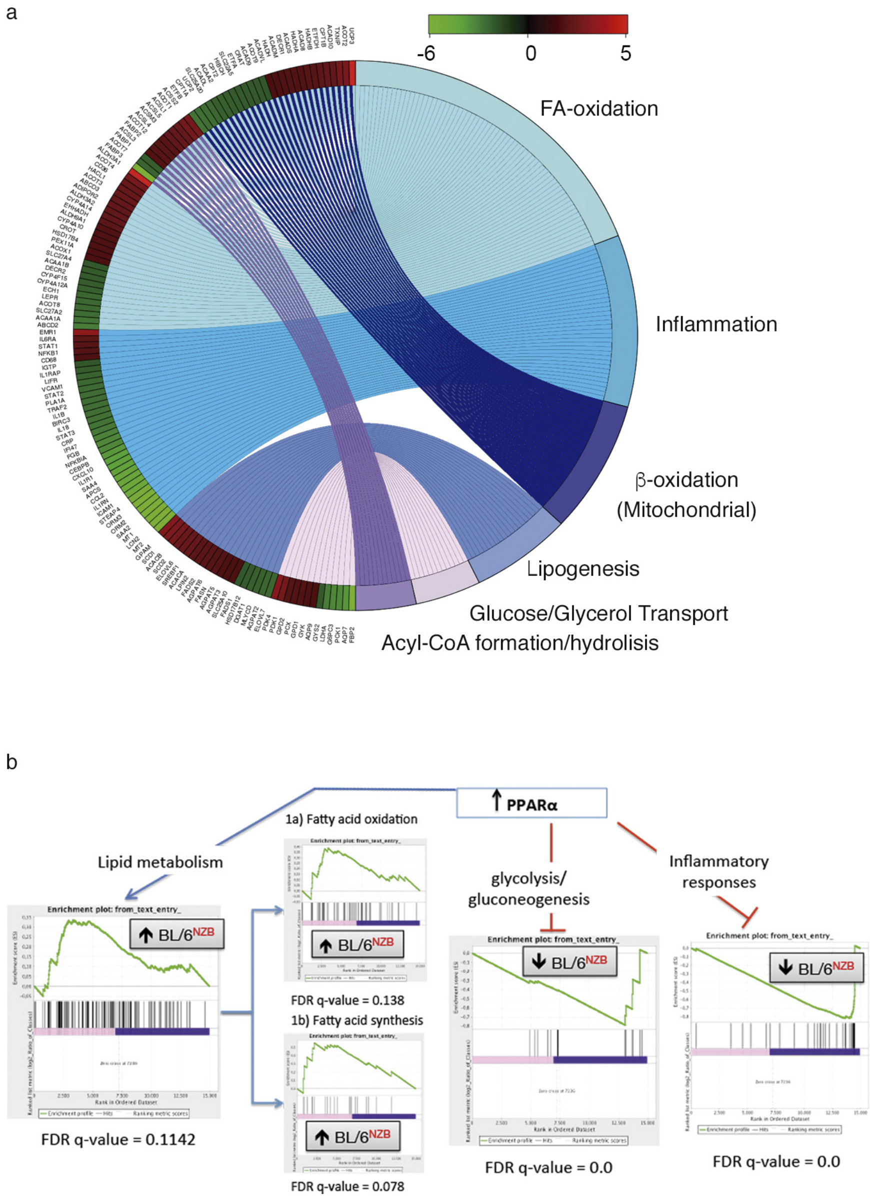
(*) p<0.05 between strains BL/6^{C57}& NZB^{NZB}

(#) p<0.05 when compared with any strain

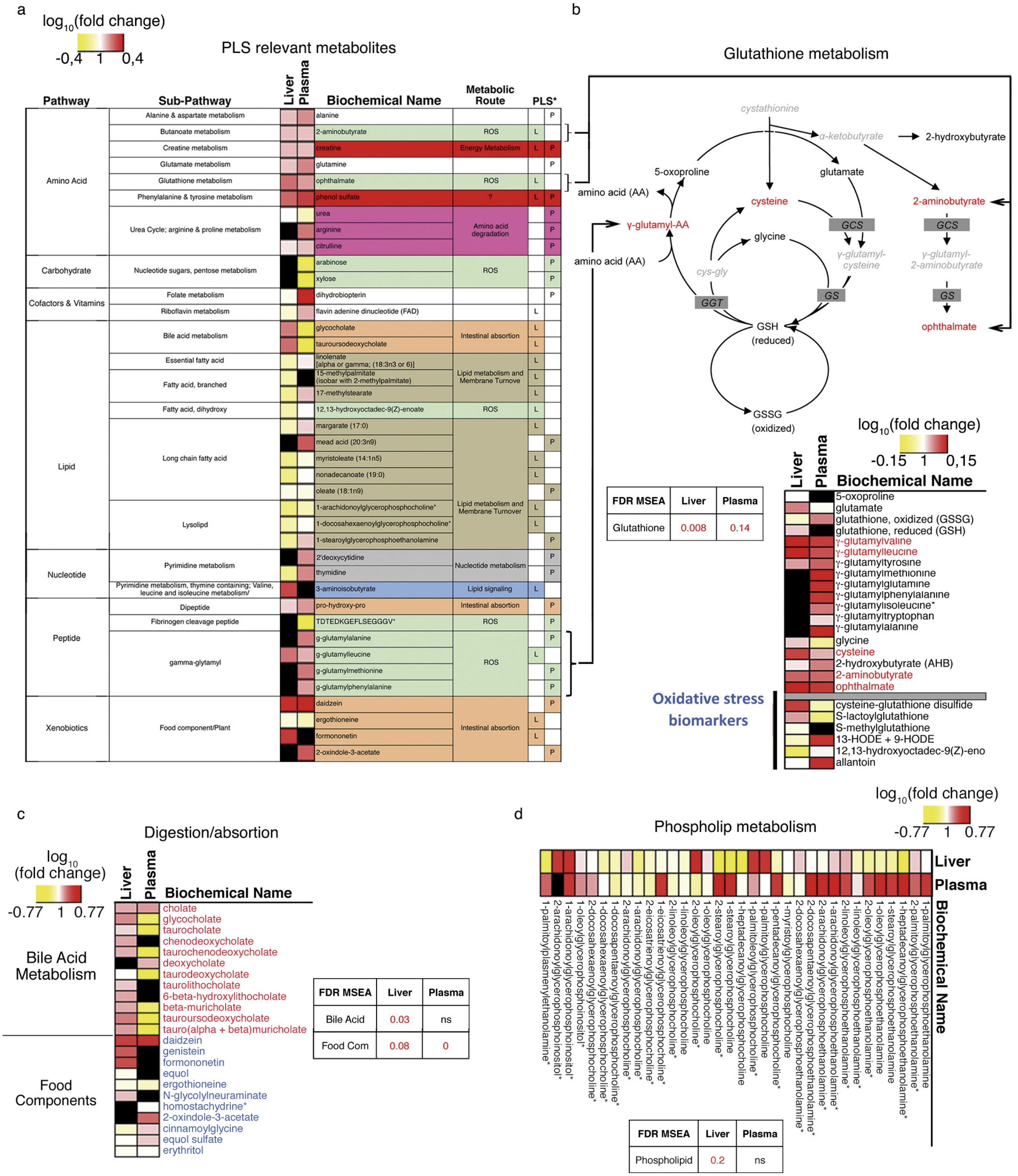
Extended Data Figure 2 | Young conplastic mouse phenotyping. **a**, Weight-gain during early life (BL/6^{NZB}, BL/6^{C57}, and NZB^{NZB} separated by sex, $n = 10$ per group). **b**, Organ-weight to body-weight ratios separated by sex ($n = 10$ per group, ANOVA test). **c**, Activity profiles (BL/6^{NZB}, BL/6^{C57}, and NZB^{NZB} separated by sex, 10-week-old mice, $n = 10$ per group). **d**, Area under the curve analysis of the activities shown in **c**, providing estimated activity intensities of the different strains. **e**, Ratios of activity to respiratory parameters in the different mouse strains. Data are means ± s.d., $*P < 0.05$, two-tailed *t*-test.



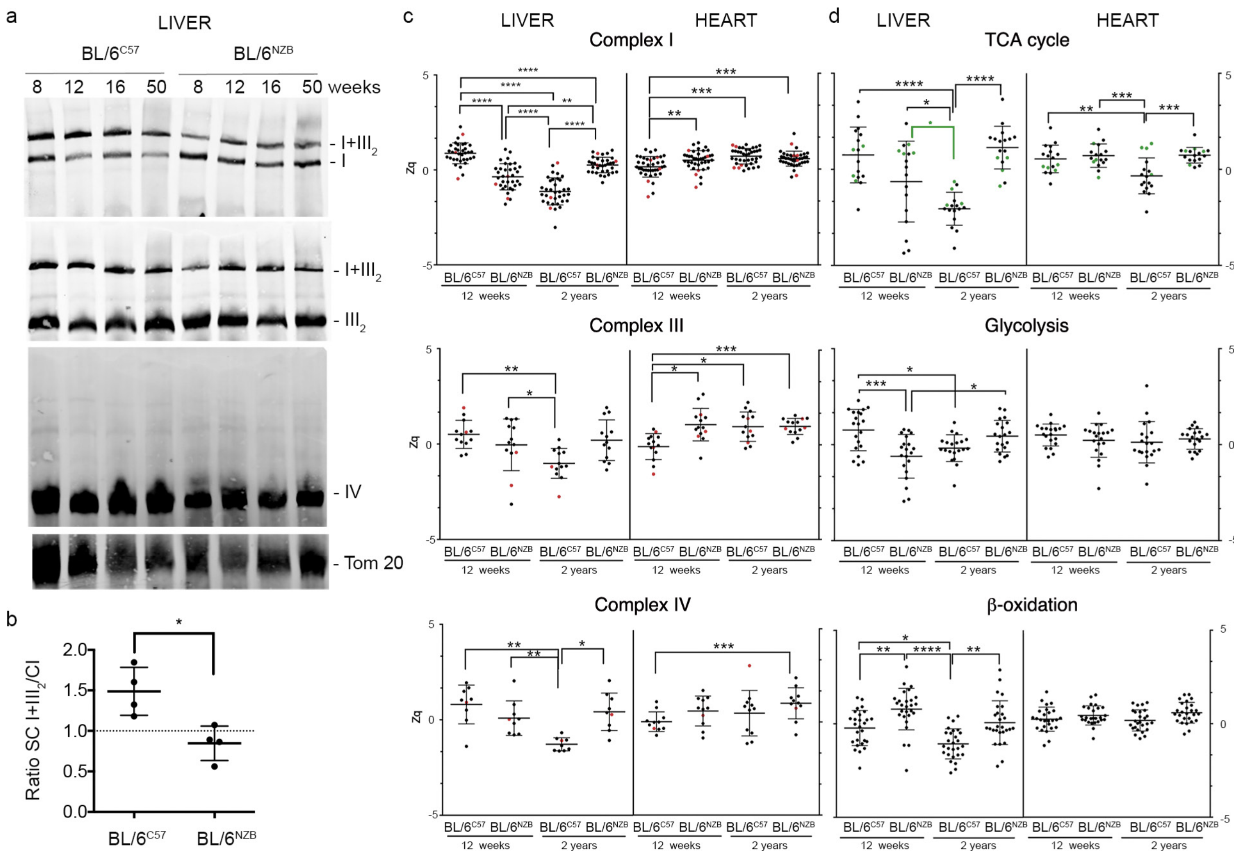
Extended Data Figure 3 | Ageing phenotype of conplastic mice. **a**, Two-year-old mice were scored for conspicuous signs of the indicated ageing parameters (0 (none) to 3 (high))[10]. **b**, Representative image of 2-year-old animals. **c**, Tumours incidence detected upon liver necropsy of animals dying from natural causes (BL/6^{NZB}, $n = 14$; BL/6^{C57}, $n = 20$ mice). **d**, Representative haematoxylin and eosin stained paraffin sections of liver adenoma, carcinoma, histiocytic sarcoma and hemangiosarcoma. **e**, Representative histological preparations of ovaries from animals of the indicated age and genotype. **f**, Lipofuscin accumulation over the reproductive period of conplastic females ($n = 5$ females per genotype and age) $*P < 0.05$, two-tailed t-test.



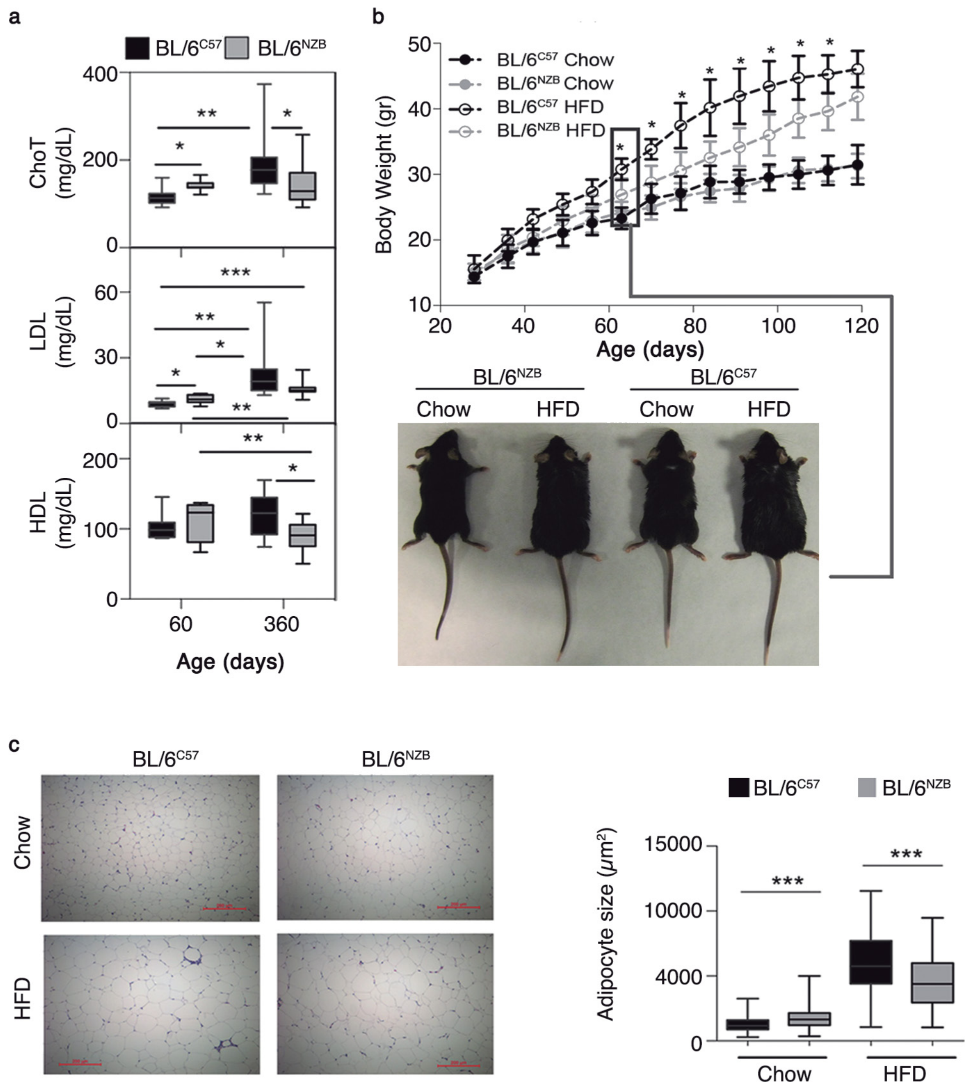
Extended Data Figure 4 | Additional transcriptomic data. a, Chord plot comparing liver expression of PPARα-regulated genes in conplastic mice. The chord plot shows gene changes (logFC) and relates them to their functional categories. **b**, Gene set enrichment analysis (GSEA), showing false discovery rate (FDR) estimates of significant and non-significant differences between conplastic mice for liver gene expression linked to the indicated processes.



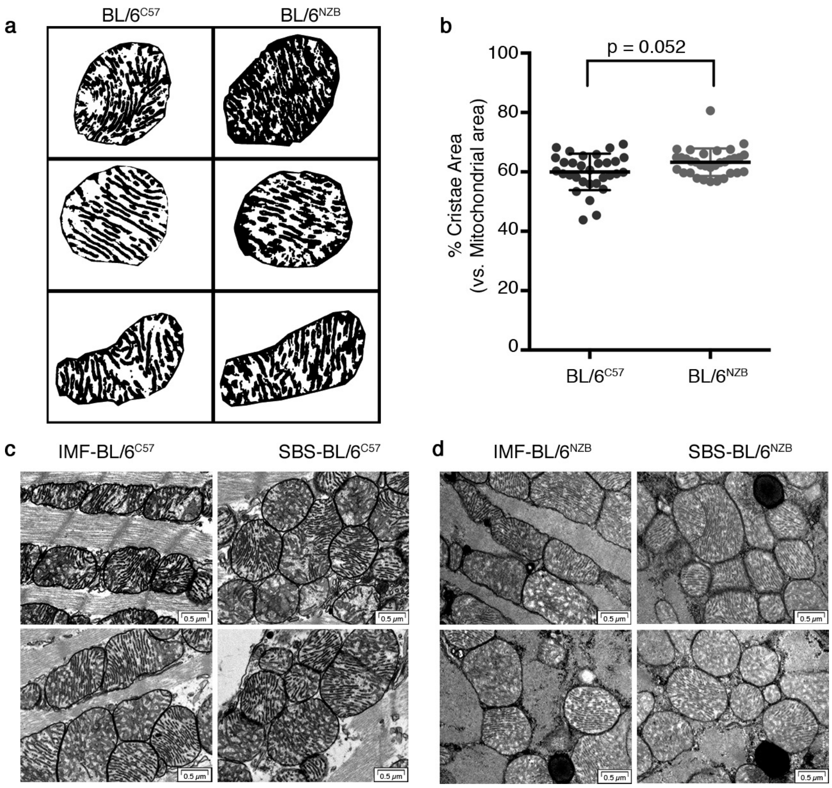
Extended Data Figure 5 | Differentially regulated metabolites and metabolic pathways in liver and plasma. a, PLS-DA-identified metabolites responsible for differential metabolic signatures (P, plasma; L, liver). The metabolite list was cross-validated, and only metabolites present in 50% or more iterations were considered. Colour highlighting indicates metabolites as potential markers of the indicated metabolic route. **b**, Glutathione metabolic pathway and heat map showing MSEA assessment of all detected metabolites implicated in GSH metabolism, regardless of whether they were identified as relevant in PLS-DA (red text indicates PLS-DA relevant), and of selected oxidative stress metabolic biomarkers. These data are also shown in Fig. 2d. **c**, Heat map of detected metabolites implicated in bile acid and xenobiotic metabolism. **d**, Heat map of detected metabolites implicated in phospholipid metabolism. Heat maps are accompanied by MSEA assessment of significant (red) or nonsignificant differences (FDR). Colour-coding in all heat maps represents the log fold difference between BL/6^{NZB} and BL/6^{C57} mice (black, not detected; white, no change (log fc = 1); red, trend for enrichment in BL/6^{NZB} samples (log fc >1); yellow, trend for depletion in BL/6^{NZB} samples (log fc <1).



Extended Data Figure 6 | Quantitative proteomics analysis of mitochondrial complexes and functional categories. **a**, Blue native electrophoresis of liver samples from 8-, 12-, 16- and 50-week-old conplastic mice. Top: complex I, (Ndufa9); middle: complex III (core-2); bottom: complex IV (Cox5b) and Tom20. **b**, Ratio of supercomplexed to free complex I, determined from the CI western blot in **a**. **c**, Relative abundance of component proteins of CI, CIV and CIII detected by iTRAQ quantitative proteomics in heart and liver of the indicated mouse strain at 12 weeks and 2 years of age. Each dot represents a single protein; red dots represent mtDNA-encoded proteins. **d**, Relative abundance of proteins of the TCA cycle, glycolysis, or β-oxidation detected by iTRAQ quantitative proteomics in heart and liver of the indicated mouse strain at 12 weeks old and 2 years of age. The four succinate dehydrogenase (complex II) subunits are highlighted in green. The iTRAQ-detected proteins and *Zq* values represented in **c** and **d** are compiled in Supplementary Data 4. Data are means ± s.d., ANOVA test. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.00001$.



Extended Data Figure 7 | Lipid metabolism phenotype. a, Serum cholesterol levels (ChoT), low-density lipoprotein (LDL), and high-density lipoprotein (HDL) in young (2 months old, $n = 8$ per genotype) and adult (12 months old, $n = 12$ per genotype). **b**, Three-month body-weight gain in males fed chow ($n = 4$ per genotype) or a HFD ($n = 8$ per genotype). **c**, Adipocyte size in epididymal fat pads of HFD-fed conplastic males ($n = = 3$ per genotype and diet: BL/6^{C57} chow diet = 368 cells, HFD = 60 cells; BL/6^{NZB} chow diet = 353 cells, HFD = 209 cells).



Extended Data Figure 8 | Mitochondrial cristae and density.
a, Representative masks of individual heart mitochondria from 12-week-old mice ($n = 3$ per genotype). **b**, The cristae area per mitochondria estimated from masks. Significance was assessed by nonparametric Mann–Whitney *U*-test. Bars and whiskers represent means ± s.d.; each point represents the percentage of cristae area in an individual mitochondria; (BL/6^{NZB}, $n = 33$; BL/6^{C57}, $n = 31$) **c**, **d**, Aspect and electro-density of subsarcolemmal (SBS) or intermyofibrillar (IMF) mitochondria from left heart ventricular apex of 12-week-old mice.

Extended Data Table 1 | Sequence differences between NZB and C57 mtDNA

Position	C57BL/6	NZB	Gen	Category	Position	C57BL/6	NZB	Gen	Category
55	G	A	tRNA Phe	non-coding	**8568**	C	T	ATP6	silent
716	A	G	rRNA 16S	non-coding	**8858**	T	C	COIII	silent
1353	A	G	rRNA 16S	non-coding	**8864**	C	T	COIII	silent
1519	G	A	rRNA 16S	non-coding	**9137**	A	G	COIII	silent
1590	G	A	rRNA 16S	non-coding	**9152**	T	C	COIII	silent
1822	T	C	rRNA 16S	non-coding	**9391**	A	G	tRNA Gly	non-coding
2201	T	C	rRNA 16S	non-coding	**9461**	T	C	ND3	silent
2340	G	A	rRNA 16S	non-coding	**9530**	C	T	ND3	silent
2525	C	T	rRNA 16S	non-coding	**9581**	C	T	ND3	silent
2766	A	G	ND1	Missense	**9599**	A	G	ND3	silent
2767	T	C	ND1	missense	**9829***	T	AA	tRNA Arg	non-coding
2798	C	T	ND1	silent	**10547**	C	T	ND4	silent
2814	T	C	ND1	silent	**10583**	A	G	ND4	silent
2840	C	T	ND1	silent	**10952**	C	A	ND4	missense
2934	C	T	ND1	missense	**11843**	G	A	ND5	silent
3194	T	C	ND1	silent	**11846**	C	T	ND5	silent
3260	A	G	ND1	silent	**11933**	A	C	ND5	silent
3422	T	C	ND1	silent	**12353**	C	T	ND5	silent
3467	T	C	ND1	silent	**12575**	T	A	ND5	silent
3599	T	C	ND1	silent	**12695**	A	G	ND5	silent
3692	A	G	ND1	silent	**12835**	T	C	ND5	missense
3932	G	A	ND2	missense	**12890**	A	G	ND5	silent
4123	C	T	ND2	silent	**13004**	G	A	ND5	silent
4276	G	A	ND2	silent	**13444**	C	T	ND5	missense
4324	T	C	ND2	silent	**13612**	T	C	ND6	silent
4408	G	A	ND2	silent	**13689**	C	T	ND6	missense
4706	A	G	ND2	missense	**13781**	A	G	ND6	missense
4732	C	T	ND2	silent	**13780**	T	C	ND6	missense
4771	T	C	ND2	silent	**13837**	A	G	ND6	silent
4885	A	C	ND2	silent	**13983**	A	G	ND6	silent
4903	T	G	ND2	silent	**14186**	T	C	CYTB	silent
5204	A	AG	tRNA Cys	non-coding	**14211**	G	A	CYTB	missense
5463	G	A	COI	missense	**14363**	A	G	CYTB	silent
5552	T	C	COI	silent	**14642**	G	A	CYTB	silent
5930	G	A	COI	silent	**14738**	C	T	CYTB	silent
6041	T	C	COI	silent	**15499**	T	A	D-Loop	intergenic
6407	C	T	COI	silent	**15549**	C	T	D-Loop	intergenic
6470	A	G	COI	silent	**15578**	A	T	D-Loop	intergenic
6575	C	T	COI	silent	**15588**	C	T	D-Loop	intergenic
6620	G	A	COI	silent	**15603**	C	T	D-Loop	intergenic
6785	G	A	COI	silent	**15657**	T	C	D-Loop	intergenic
7411	A	G	COII	silent	**15916**	C	T	D-Loop	intergenic
7870	G	A	ATP8	silent	**16017**	A	C	D-Loop	intergenic
8439	A	G	ATP6	silent	**16268**	A	G	D-Loop	intergenic
8467	T	C	ATP6	silent	**16272**	T	C	D-Loop	intergeni

Differences were identified by mtDNA-NGS. The *9829insAA position was confirmed by Sanger sequencing.

Extended Data Table 2 | Analysis of the mtDNA mutation load

Mouse type	Animal	Mutations	Average base coverage depth	Uniformity of base Coverage
BL/6^{C57}_12w	1	0	8769	99.52%
BL/6^{C57}_12w	2	0	8910	99.19%
BL/6^{C57}_12w	3	1	6289	99.08%
BL/6^{C57}_12w	4	1	9766	99.45%
BL/6^{C57}_12w	5	0	8923	99.28%
Total		**2**		
BL/6^{C57}_40w	1	0	8806	98.96%
BL/6^{C57}_40w	2	0	8884	99.38%
BL/6^{C57}_40w	3	0	8466	99.30%
BL/6^{C57}_40w	4	2	9451	99.39%
Total		**2**		
BL/6^{C57}_80w	1	0	8728	99.44%
BL/6^{C57}_80w	2	1	8857	99.29%
BL/6^{C57}_80w	3	0	9710	99.49%
BL/6^{C57}_80w	4	1	11139	99.35%
BL/6^{C57}_80w	5	0	9833	99.46%
Total		**2**		
BL/6^{NZB}_12w	1	1	10992	99.54%
BL/6^{NZB}_12w	2	0	9237	99.45%
BL/6^{NZB}_12w	3	0	8417	99.49%
BL/6^{NZB}_12w	4	0	8341	99.44%
BL/6^{NZB}_12w	5	0	9672	99.64%
Total		**1**		
BL/6^{NZB}_40w	1	0	8898	99.52%
BL/6^{NZB}_40w	2	1	8680	99.49%
BL/6^{NZB}_40w	3	0	8090	99.60%
BL/6^{NZB}_40w	4	0	8109	99.43%
Total		**1**		
BL/6^{NZB}_80w	1	0	9495	99.39%
BL/6^{NZB}_80w	2	0	8288	99.33%
BL/6^{NZB}_80w	3	0	10994	99.53%
BL/6^{NZB}_80w	4	0	8799	99.58%
BL/6^{NZB}_80w	5	0	9759	99.39%
Total		**0**		

Mouse type	Animal	Position	R/A	AA	FDP	%Het	Gen	Category
BL/6^{C57}_12w	3	8258	G/T	G/V	9962	2,3	ATP6	Missense
BL/6^{C57}_12w	4	5636	A/T	W/C	9946	4,7	COXI	Missense
BL/6^{C57}_40w	4	1025	C/T	none	9977	7,0	tRNA^{Val}	Non-coding
BL/6^{C57}_40w	4	8637	G/A	V/I	9943	28,5	COXIII	Missense
BL/6^{C57}_80w	2	80	T/G	none	8660	2,4	12S rRNA	Non-coding
BL/6^{C57}_80w	4	18	A/G	none	6205	47,5	tRNA^{Phe}	Non-coding
BL/6^{NZB}_12w	1	7633	G/A	none	9979	3,4	COXII	Silent
BL/6^{NZB}_40w	2	10624	C/A	none	7931	5,8	ND4	Silent

doi:10.1038/nature18643

Sliding sleeves of XRCC4–XLF bridge DNA and connect fragments of broken DNA

Ineke Brouwer[1]*, Gerrit Sitters[1]*†, Andrea Candelli[1]†, Stephanie J. Heerema[1]†, Iddo Heller[1], Abinadabe J. de Melo[2], Hongshan Zhang[2], Davide Normanno[2], Mauro Modesti[2]§, Erwin J. G. Peterman[1]§ & Gijs J. L. Wuite[1]§

Non-homologous end joining (NHEJ) is the primary pathway for repairing DNA double-strand breaks (DSBs) in mammalian cells[1]. Such breaks are formed, for example, during gene-segment rearrangements in the adaptive immune system or by cancer therapeutic agents. Although the core components of the NHEJ machinery are known, it has remained difficult to assess the specific roles of these components and the dynamics of bringing and holding the fragments of broken DNA together. The structurally similar XRCC4 and XLF proteins are proposed to assemble as highly dynamic filaments at (or near) DSBs[2]. Here we show, using dual- and quadruple-trap optical tweezers combined with fluorescence microscopy, how human XRCC4, XLF and XRCC4–XLF complexes interact with DNA in real time. We find that XLF stimulates the binding of XRCC4 to DNA, forming heteromeric complexes that diffuse swiftly along the DNA. Moreover, we find that XRCC4–XLF complexes robustly bridge two independent DNA molecules and that these bridges are able to slide along the DNA. These observations suggest that XRCC4–XLF complexes form mobile sleeve-like structures around DNA that can reconnect the broken ends very rapidly and hold them together. Understanding the dynamics and regulation of this mechanism will lead to clarification of how NHEJ proteins are involved in generating chromosomal translocations[3,4].

It has long been assumed that the NHEJ components assemble sequentially, starting with the Ku heterodimer (Ku70 and Ku80 subunits) binding to DNA ends and protecting them, followed by the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), which promotes end synapsis. According to this model, end-processing factors are recruited next, and DNA ligase 4 (LIG4) in complex with XRCC4 stimulated by XLF ligates the broken ends. Recent studies have, however, challenged this sequence of events[5]. In particular, XRCC4 is present in large excess over LIG4 (ref. 6) and associates with XLF to form complexes promoting LIG4-independent broken end bridging[7–9]. Crystal structures and biochemical analysis of XRCC4–XLF complexes revealed that they form alternating helical filaments capable of bridging DNA[10–13]. The biological relevance of these filaments is the subject of current studies. Recent super-resolution microscopy of human U2OS cells showed that XRCC4, XLF and LIG4 assemble as filamentous structures adjacent to bleomycin-induced DSBs[2]. XRCC4 and XLF can form similar structures, even in cells lacking LIG4, indicating that XRCC4–XLF filaments have a central role in NHEJ. To gain insight into the dynamics of these filaments and their interactions with DNA, we used a single-molecule approach combining dual optical trapping, microfluidics and fluorescence microscopy[14–16]. To study DNA bridging, we needed to extend this approach by adding two additional traps to our instrument, thus allowing the independent manipulation of two double-stranded DNA (dsDNA) molecules and the direct visualization of interactions between them. This instrument allowed us to investigate the bringing and holding together of DNA fragments by human XRCC4 and XLF, key intermediate steps in NHEJ.

First, we studied the interaction of fluorescently labelled XLF and XRCC4 (Extended Data Figs 1–3) on DNA held taut by two optically trapped microspheres (Fig. 1). Such experiments show that both proteins bind to end-occluded dsDNA but not to stretched

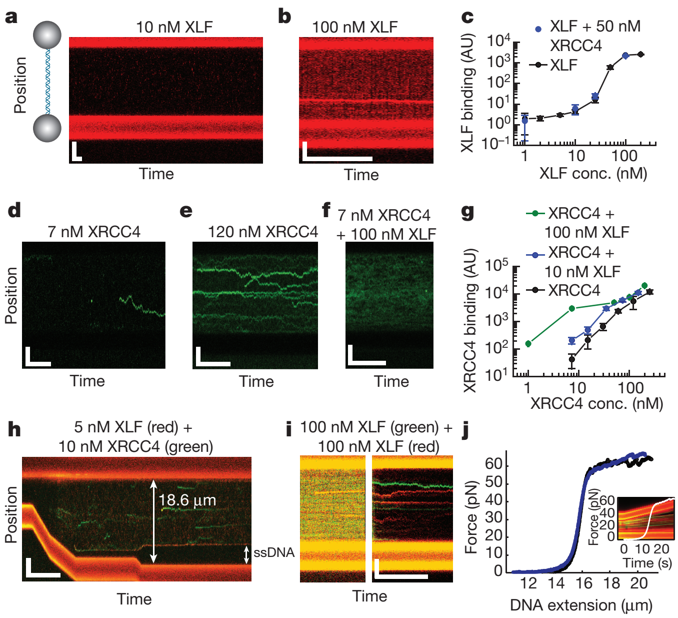


Figure 1 | Interaction of XRCC4 and XLF with DNA. a, b, d–f, Kymographs of dsDNA with 10 nM XLF (3 ms pixel integration time) (**a**), 100 nM XLF (**b**), 7 nM XRCC4 (**d**), 120 nM XRCC4 (**e**) and 7 nM XRCC4 and 100 nM XLF (only XRRC4 was excited) (**f**). **c**, **g**, Concentration-dependent DNA-binding of XLF (**c**) and XRCC4 (**g**). AU, arbitrary units. Error bars denote s.e.m. **h**, Kymograph of XRCC4 and XLF binding. At tensions above ~65 pN the DNA overstretches by unpeeling from DNA ends until 18% of the molecule consists of ssDNA. No binding of XLF and XRCC4 to ssDNA was observed. **i**, Kymographs of XRCC4 and XLF binding in the presence of proteins (left) and after exchange with protein-free buffer (right; same construct and settings, 6 s later). **j**, Force-extension curve of dsDNA incubated (under tension) in 200 nM XRCC4 and 200 nM XLF (blue) compared to bare dsDNA (black). Inset, simultaneous kymograph. Dark regions in the DNA appearing above ~60 pN are due to ssDNA generated by force-induced melting. Red, enhanced green fluorescent protein (eGFP)-labelled XLF; green, XRCC4–Alexa Fluor 555; yellow, colocalization. Scale bars, 4 μm and 5 s. Pixel integration time: 3 ms (**a**) or 0.1 ms (others). Data are representative of 13 (**a**), 17 (**b**), 15 (**d**), 11 (**e**), 10 (**f**), 6 (**h**), 12 (**i**) and 8 (**j**) experiments.

[1]Department of Physics and Astronomy and LaserLab, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands. [2]Cancer Research Center of Marseille, CNRS UMR7258, Inserm U1068, Institut Paoli-Calmettes, Aix-Marseille Université UM105, 13273 Marseilles, France. †Present addresses: LUMICKS B.V., 1081 HV Amsterdam, The Netherlands (G.S., A.C.); Kavli Institute of Nanoscience, Delft University of Technology, 2628 CJ Delft, The Netherlands (S.J.H.).
*These authors contributed equally to this work.
§These authors jointly supervised this work.

single-stranded DNA (ssDNA) (Fig. 1h and Extended Data Fig. 4a). In the absence of XRCC4, XLF binds to DNA transiently at low concentrations (Fig. 1a), and forms stable oligomers at higher concentrations (Fig. 1b), a large fraction of which is able to diffuse over DNA. The amount of XLF bound to the DNA increased drastically with increasing XLF concentrations and saturated around 100 nM (Fig. 1c), suggesting cooperative binding. XRCC4 exhibits a different behaviour: at all concentrations tested, XRCC4 binds DNA in a less cooperative manner, forming stable oligomers that are able to diffuse over DNA (Fig. 1d, e, g). Both XLF and XRCC4 oligomers appeared to bind directly from solution and, once bound to the DNA, oligomers did not appear to grow in size (Extended Data Fig. 4e).

Next, we addressed the interaction between the two proteins. In the presence of XLF, the affinity of XRCC4 for the DNA was enhanced (Fig. 1f, h) and XRCC4 was distributed more homogenously over the DNA (compare Fig. 1d to 1f). By contrast, XRCC4 did not significantly alter the affinity of XLF for DNA (Fig. 1c). We noticed that when a DNA molecule was moved to a protein-free solution after incubation in the presence of XLF and XRCC4, the fluorescence signal decreased much faster than the estimated bleaching rate (Fig. 1i), indicating that proteins (presumably dimers) were dissociating rapidly (that is, within 10 ms) and thus were bound transiently to the DNA (Extended Data Fig. 4d). The proteins that remain are assembled in oligomers (Extended Data Fig. 4b, c), most of which diffuse (88 out of 94). Moreover, these protein complexes dissociate at least four orders of magnitude slower (tens of seconds to tens of minutes) than the transiently bound dimers (Extended Data Figs 4f and 5g, h). Of these oligomers, $29 \pm 5\%$ (mean $\pm$ s.e.m.) contained both XRCC4 and XLF, $9 \pm 3\%$ only XRCC4 and $62 \pm 8\%$ only XLF (of a total of 94 detected oligomers at a concentration of 50 nM each). Together, these results suggest that XLF regulates the loading of (heteromeric) complexes onto DNA, possibly by influencing the oligomeric state of XRCC4 in solution[17,18].

A large fraction of the stably bound complexes containing XLF, XRCC4 or both switched back and forth between static and diffusive binding (36 out of 94) (Extended Data Fig. 6a). A smaller fraction seemed to be fixed (>120 s) on the DNA (4 out of 94), illustrating that the complexes can interact with the DNA in different ways. The typical switching time between the static and the mobile state is approximately 10 s (Extended Data Fig. 6f). For the mobile fraction, diffusion coefficients were $0.7 \pm 0.1\,\mu m^2\,s^{-1}$, $0.7 \pm 0.1\,\mu m^2\,s^{-1}$ and $0.6 \pm 0.2\,\mu m^2\,s^{-1}$ for XLF, XRCC4 and XRCC4–XLF complexes, respectively, at 25 mM KCl (Extended Data Fig. 6b). These values were not strongly dependent on DNA tension and complex size, but increased at higher ionic strength (Extended Data Fig. 6c, d). Under all conditions tested, the diffusion was faster than would be expected for complexes undergoing rotation-coupled diffusion along the DNA[19] (Extended Data Fig. 6e), suggesting sliding without tracking the helical pitch. Furthermore, the mechanical properties of DNA were not altered upon binding of XRCC4 and XLF, even at high density (Fig. 1j). On the basis of these observations, we propose that, in mobile complexes, the proteins bind around the DNA as a sleeve: once loaded, the stably bound complexes slide along the DNA in a diffusive manner experiencing little friction.

Next, we addressed whether these XRCC4–XLF complexes have a role in bringing and holding DNA fragments together, crucial steps in NHEJ. To study this in a well-controlled way (Extended Data Fig. 7), we developed an instrument combining quadruple-trap optical tweezers with wide-field fluorescence microscopy. Using this instrument, we independently trapped two end-occluded dsDNA molecules, held them taut and wrapped them (Extended Data Fig. 8). We incubated these intertwined dsDNA molecules in a channel containing 200 nM XLF and 200 nM XRCC4 for 2 min. We moved these DNA molecules into a protein-free buffer, unwrapped them and visualized whether bridge formation had occurred using fluorescence. The mechanics of the bridge were probed by moving the microspheres and detecting the

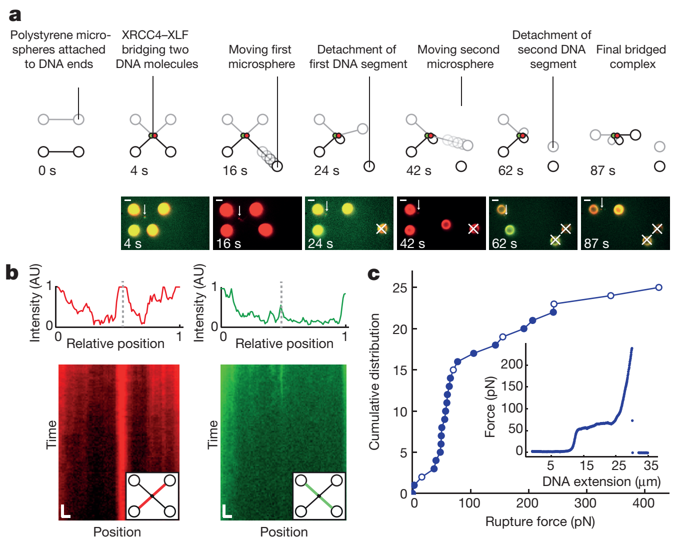


Figure 2 | DNA bridging by XRCC4 and XLF. a, Fluorescence images (bottom) and corresponding cartoons (top) of two bridged dsDNA molecules after wrapping, incubating (2 min in 200 nM XRCC4 and 200 nM XLF) and unwrapping. At 4, 24, 62 and 87 s images are overlays of green and red images. Arrows denote bridge location. X marks microsphere detached from bridge. **b**, Intensity profiles (top) and logarithmically scaled kymographs (bottom) along DNA indicated in insets. Dashed line denotes bridge position. **c**, Distribution of rupture forces of bridged complexes (such as shown at 87 s in **a**). Open circles measured in 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate and 100 μg ml^{-1} BSA (pH 7.9), filled circles in 20 mM Tris-HCl (pH 7.6), 25 mM KCl and 1 mM DTT. Inset, force-extension curve of such a complex. Green, XRCC4–Alexa Fluor 555, red, eGFP–XLF, yellow, colocalization. Scale bars, 5 s and 1 μm. Data are representative examples of 117 experiments.

resulting forces. An example of such an experiment is shown in Fig. 2a (see also Supplementary Video 1 and Extended Data Fig. 9). In this experiment, the fluorescence image showed that a bridge was present after incubating and unwrapping the DNA molecules (Fig. 2a, 4 s). At the DNA–DNA interaction site, a yellow patch is clearly visible, demonstrating that both XRCC4 and XLF are present at the bridge. Subsequently, we moved the bottom-right microsphere to apply a high force (>100 pN) until we observed a sudden force drop, due to one of the DNA ends detaching from the bottom-right bead (Fig. 2a, 24 s). Thereafter, the top-right microsphere was moved, resulting in the detachment of a second DNA end from this microsphere (Fig. 2a, 62 s). The resulting tether now consisted of two DNA fragments, originating from different DNA molecules, held together by XRCC4–XLF. The mechanical properties of this protein–DNA complex were indistinguishable from those of bare dsDNA, with a different length from the original DNA molecules. This experiment directly demonstrates that XRCC4–XLF protein complexes can bridge and hold together broken DNA fragments.

To quantify this bridging behaviour, we repeated the dual-DNA experiment 203 times. From these experiments, several key features of bridging were extracted. In contrast to the observation on single DNA molecules that only ~30% of the complexes contained both XRCC4 and XLF, all 152 bridges between two DNA molecules contained both proteins (Extended Data Fig. 10). Bridging was observed in two-thirds of the cases, both for wrapped (140 out of 183 tests) and crossed (12 out of 20 tests) DNA pairs (Extended Data Fig. 8). Bridges also formed between DNA pairs that were pre-incubated with proteins and subsequently wrapped/unwrapped in the absence of protein (Extended Data Fig. 5a, b). XRCC4–XLF bridges were stable for at least 90 min at low force (Extended Data Fig. 5g, h). High force applied to the bridge resulted in rupture: in 9 out of 96 cases the protein bridge ruptured (Extended Data Fig. 5e, f), and in the remainder of the cases either

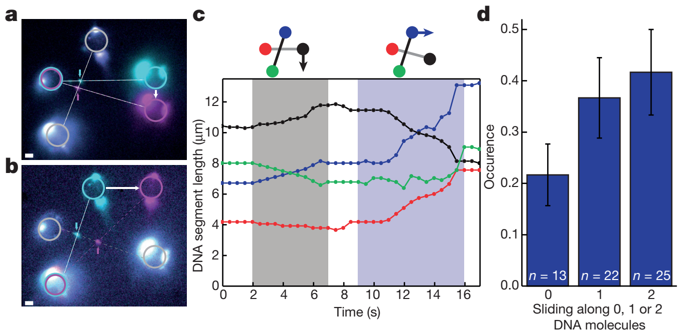


Figure 3 | Mobility of XRCC4–XLF bridges. a, **b**, Overlays of eGFP–XLF logarithmically scaled fluorescence before (cyan) and after (magenta) moving a microsphere (white arrow). DNA molecules were incubated in crossed configuration. Circles denote microsphere positions; dashed lines denote DNA; coloured arrows denote bridge location. Data are representative examples of 13 experiments. **c**, Distance from bridge to bead edges (of complex shown in **a** and **b**) as a function of time. Shaded regions denote bead motion as indicated by the arrows in the cartoons. **d**, Histogram of XRCC4–XLF bridge mobility (both wrapped and crossed configurations). Error bars denote s.e.m.

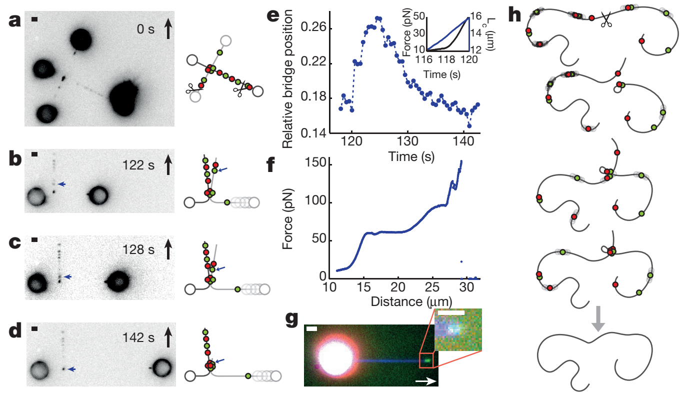


Figure 4 | XRCC4–XLF complexes keep DNA fragments together. a–**d**, Inverse fluorescence images of XRCC4–Alexa Fluor 555 eGFP–XLF bridge (see Supplementary Video 3; 532 nm excitation). DSBs are introduced using SfoI, cleaving each DNA in one site. Free ends extend in buffer flow (black arrows). Blue arrows denote location of one XRCC4–XLF complex. **e**, Relative position of bridge shown in **b**–**d** between the beads. Inset denotes contour length and force as a function of time. **f**, Force-extension curve of the complex in **b**–**e**. **g**, Fluorescence image of proteins at the end of a 32.2-kb flow-stretched blunt-ended SYTOX Red-stained DNA molecule. Green, XRCC4–Alexa Fluor 555; red, eGFP–XLF; blue, SYTOX Red; white, colocalization. Arrow shows flow direction. Inset, zoom with different intensity scaling. **h**, Proposed mechanism for XRCC4 (green) and XLF (red) in bridging and holding together DNA fragments (grey) after a DSB (scissors). Scale bars, 2 µm. Data are representative examples of 30 (**a**–**f**) and 14 (**g**) experiments.

one of the DNA molecules broke or detached from a microsphere (Fig. 2a, 24 s). To assess the strength of the bridges more directly, we determined the rupture forces of bridged DNA fragments tethered between two microspheres (for example, the tether in Fig. 2a, 62–87 s). Such complexes ruptured over a wide range of forces (Fig. 2c). Most events occurred during DNA overstretching (at ~65 pN), indicating breakage of the DNA due to nicks[20] and therefore not rupture of the protein bridges. A substantial fraction broke at higher forces, up to 250 pN, by detachment of the biotin–streptavidin DNA–microsphere linkers[21], by rupture of the protein bridge, or by the DNA sliding out of the sleeve. Together, these experiments indicate that XRCC4–XLF complexes readily form highly stable and strong (at least >65 pN) DNA–protein bridges.

To relate our observations of the stable XRCC4–XLF bridges between two DNA molecules to the highly mobile sleeves present on single DNA molecules, we tested whether the bridges are also mobile. In our quadruple-trap assay, we moved one DNA molecule with respect to the other and tracked the position of the protein bridge in the fluorescence images (for example, Fig. 3a–c, Extended Data Fig. 9 and Supplementary Video 2). In 47 out of 60 such experiments, we observed that bridges could slide along one (22 cases) or both (25 cases) of the DNA molecules (Fig. 3d). This mobility indicates that the bridges consist of one or a few sleeves. A single sleeve could surround two DNA molecules, allowing them to slide independently within the sleeve. Alternatively, two mobile sleeves, each surrounding a single DNA molecule, might interact to form a bridge. The static binding mode observed on single DNA molecules (Extended Data Fig. 6a) might explain why some of the bridges cannot slide along one or both DNA molecules.

The observations so far, using end-occluded DNA, raise the question of how XRCC4–XLF complexes interact with DNA fragments and DNA ends. Fast bridge formation between the two fragments is vital for NHEJ since it avoids interactions that could lead to chromosomal translocations. To study the interaction of XRCC4–XLF with DSBs, XRCC4–XLF bridges ($n = 30$) were exposed to the restriction enzyme SfoI, inducing a single DSB in both DNA molecules. Figure 4a–d illustrates such an experiment (see also Supplementary Video 3). This experiment was performed in slight buffer flow, such that, after introducing DSBs, the DNA ends were extended in the direction of the flow to prevent the DNA from bridging onto itself. XRCC4–XLF bridges are most of the time still capable of sliding along the DNA fragments (Fig. 4e). At some point, this sliding is stalled, probably when the bridge runs into a stationary XLF and/or XRCC4 complex or a DNA end. Typically, we do not see the DNA ends sliding out of the bridge when the bridge reaches the end of the DNA fragment (Fig. 4f). These experiments show directly that after DSB formation, XRCC4–XLF bridges are able to hold together the broken DNA fragments such that their ends remain in close proximity to each other, while the mobility of the complexes ensures that the ends remain accessible for further processing by other repair factors.

To study more directly what happens when an XRCC4–XLF complex encounters a free DNA end, we tethered a blunt-ended dsDNA molecule to a single microsphere and extended it using buffer flow. We expected that DNA-bound XRCC4–XLF complexes are driven towards the DNA end by the buffer flow. Indeed, in fluorescence images recorded after moving the construct to a protein-free environment, we observed accumulation of XRCC4 and XLF at the DNA end (Fig. 4g) in all 14 experiments. This suggests that some of the protein sleeves do not seem to slide off the blunt end under the influence of buffer flow. However, further experimentation is required to understand how sleeves intrinsically interact with ends and how they are affected by other factors, such as Ku.

In this study we have demonstrated that the combination of dual- and quadruple-trap optical tweezers with fluorescence microscopy is a powerful tool to study the roles of XRCC4 and XLF in NHEJ. Our experiments show that binding of XRCC4 is stimulated by XLF and that together they form stably bound, yet mobile complexes on dsDNA. These complexes can form strong, mobile DNA bridges capable of holding the DNA fragments in close proximity to one other. Our and previous[2] observations can be explained as follows (Fig. 4h): XRCC4 and XLF bind to the DNA fragments of a DSB in a highly diffusive manner. These complexes can bridge the DNA fragments at any location, preventing the fragments from moving apart. The mobility of the bridges could allow them to slide along the DNA leaving the ends accessible for processing. Our method with the capability to observe directly how XRCC4–XLF complexes dynamically interact with DNA and each other in NHEJ can be used in future experiments to address remaining questions in NHEJ, including how the loading of XRCC4–XLF complexes is restricted to broken DNA, how they are remodelled and regulated, and how Ku, DNA-PKcs, LIG4 and other NHEJ factors come into play.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 April; accepted 13 June 2016.
Published online 20 July 2016.

1. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79,** 181–211 (2010).
2. Reid, D. A. *et al.* Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proc. Natl Acad. Sci. USA* **112,** E2575–E2584 (2015).
3. Ghezraoui, H. *et al.* Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol. Cell* **55,** 829–842 (2015).
4. Gelot, C. *et al.* The cohesin complex prevents the end joining of distant DNA double-strand ends. *Mol. Cell* **61,** 15–26 (2016).
5. Cottarel, J. *et al.* A noncatalytic function of the ligation complex during nonhomologous end joining. *J. Cell Biol.* **200,** 173–186 (2013).
6. Mani, R. S. *et al.* Dual modes of interaction between XRCC4 and polynucleotide kinase/phosphatase: implications for nonhomologous end joining. *J. Biol. Chem.* **285,** 37619–37629 (2010).
7. Graham, T. G. W., Walter, J. C. & Loparo, J. J. Two-stage synapsis of DNA ends during non-homologous end joining. *Mol. Cell* **61,** 850–858 (2016).
8. Roy, S. *et al.* XRCC4's interaction with XLF is required for coding (but not signal) end joining. *Nucleic Acids Res.* **40,** 1684–1694 (2012).
9. Roy, S. *et al.* XRCC4/XLF interaction is variably required for DNA repair and is not required for ligase IV stimulation. *Mol. Cell. Biol.* **35,** 3017–3028 (2015).
10. Ropars, V. *et al.* Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proc. Natl Acad. Sci. USA* **108,** 12663–12668 (2011).
11. Hammel, M. *et al.* XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. *J. Biol. Chem.* **286,** 32638–32650 (2011).
12. Wu, Q. *et al.* Non-homologous end-joining partners in a helical dance: structural studies of XLF–XRCC4 interactions. *Biochem. Soc. Trans.* **39,** 1387–1392 (2011).
13. Andres, S. N. *et al.* A human XRCC4–XLF complex bridges DNA. *Nucleic Acids Res.* **40,** 1868–1878 (2012).
14. Heller, I., Hoekstra, T. P., King, G. A., Peterman, E. J. & Wuite, G. J. Optical tweezers analysis of DNA-protein complexes. *Chem. Rev.* **114,** 3087–3119 (2014).
15. Candelli, A., Wuite, G. J. L. & Peterman, E. J. G. Combining optical trapping, fluorescence microscopy and micro-fluidics for single molecule studies of DNA-protein interactions. *Phys. Chem. Chem. Phys.* **13,** 7263–7272 (2011).
16. Heller, I. *et al.* STED nanoscopy combined with optical tweezers reveals protein dynamics on densely covered DNA. *Nat. Methods* **10,** 910–916 (2013).
17. Hammel, M., Yu, Y., Fang, S., Lees-Miller, S. P. & Tainer, J. A. XLF regulates filament architecture of the XRCC4·ligase IV complex. *Structure* **18,** 1431–1442 (2010).
18. Mahaney, B. L., Hammel, M., Meek, K., Tainer, J. A. & Lees-Miller, S. P. XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair. *Biochem. Cell Biol.* **91,** 31–41 (2013).
19. Blainey, P. C. *et al.* Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.* **16,** 1224–1229 (2009).
20. van Mameren, J. *et al.* Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proc. Natl Acad. Sci. USA* **106,** 18231–18236 (2009).
21. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* **397,** 50–53 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Fourquet and S. Audebert from the CRCM Proteomics facility for help with MS analysis and G. King and A. S. Biebricher for help with the optical tweezers instruments. This work was supported by VICI grants (G.J.L.W. and E.J.G.P.), a VENI grant (I.H.) and a TopTalent grant (A.C.) from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, a European Research Council starting grant (G.J.L.W.), the French National Cancer Institute (grant PLBIO13-103) (M.M.), the ARC Foundation for Cancer Research (M.M.), the A*MIDEX project (no. ANR-11-IDEX-0001-02), the 'Investissements d'Avenir' French Government program (M.M.), fellowship no. 0558/12-5 from the Brazilian program for Coordination for the Improvement of Higher Education Personnel (A.J.M.), and a fellowship from the Collège of Aix-Marseille Université (H.Z.)

Author Contributions M.M., E.J.G.P. and A.C. conceived the study. I.B., G.S., A.C. and S.J.H. performed the single-molecule experiments. I.H. built the quadruple-trap instrument and advised on the force-fluorescence experiments. A.J.M., H.Z., D.N. and M.M. purified the proteins and performed the biochemical analysis of the protein samples. I.B., G.S., M.M., E.J.G.P. and G.J.L.W. wrote the manuscript. E.J.G.P., M.M. and G.J.L.W. led the research, the analysis and the interpretation of the results. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. (mauro.modesti@inserm.fr) or G.J.L.W. (g.j.l.wuite@vu.nl).

Reviewer Information *Nature* thanks M. Morse, M. Nabuan Naufer, M. Williams and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Dual-trap optical tweezers with confocal fluorescence. The optical set-up used for dual-trap optical trapping experiments has been described previously[16]. In brief, two optical traps are generated using a 10 W 1,064 nm CW fibre laser (YLR-10-LP, IPG Photonics), of which the beam is expanded using a 75 mm and 150 mm lens. To align the trapping beam with the confocal DNA imaging, the collimation can be modified by translating the 75 mm lens on an automated stage (AG-LS25). The trapping beam is split into two independently steerable paths using a polarizing beam splitter. In one trap, steering is achieved with an accurate piezo mirror (Nano-MTA2X Aluminium, Mad City Labs) and in the other trap with a coarse positioning piezo stepper mirror (AG-M100N). After recombining the two paths they are coupled into the objective via two 300 mm lenses. Back-focal plane interferometry is used for force measurements and the microsphere positions are determined using template-matching on images obtained from bright field LED illumination on a metal-oxide semiconductor (CMOS) camera (DCC1545M, Thorlabs). For the confocal fluorescence excitation two excitation bands (centred at 467 nm and 543 nm) from a single laser system (ALP-710-745-SC, Fianium Ltd) are used. We selected these bands using an AOTF (AOTFnc-VIS-TN, AA Opto-Electronic). For beam scanning, a fast tip/tilt piezo mirror (S-334.1SD, Physik Instrumente GmbH & Co, scan rate up to 200 Hz) is used. For detection, fibre-coupled APDs (APDs SPCM-AQRH-14-FC, fibres SPCM-QC9, Perkin Elmer) are used. The multimode fibres (62 μm diameter) serve as pinholes that provide the background rejection.

Quadruple-trap optical tweezers with wide field fluorescence. The setup used for dual DNA manipulation using quadruple-trap optical tweezers has been demonstrated previously[22]. In brief, four optical traps are created using a 20 W 1064 nm CW fibre laser (YLR-20-LP-IPG, IPG Photonics) with a coupled optical isolator. The laser beam is split in two paths by a polarizing beam splitter after which they are expanded using lenses with focal lengths of 75 mm and 150 mm. Both beams are split again, one by a 50/50 beam splitter and one by a polarizing beam splitter after the polarization is rotated 45° by a half-wave plate, creating four optical paths of which three have the same polarization. The 150 mm lenses are placed on a linear stage (AG-LS25, Newport) to modify the collimation and with that the height of the traps, enabling manipulation of 2 of the traps in 3 dimensions. Beam steering of three traps is done by course-positioning piezo stepper mirrors (AG-M100N) and one is steered by an accurate piezo mirror (Nano-MTA2X Aluminium, Mad City Labs). The four optical paths are recombined into two paths using a 50/50 beam splitter cube (in this way, 50% of the power is lost) and a polarizing beam-splitter cube, after which these two paths are recombined using 50/50 beam splitter cube (here, again 50% of the laser power is lost). Two 300 mm lenses are used to couple the beam into the objective. Force detection on two trapped microspheres is achieved on two position sensitive detectors (PSDs, DL100-7PCBA3, Pacific Silicon Sensor) by back focal plane interferometry. To detect the force on the first trapped microsphere, three paths with the same polarization are separated and dumped by a polarizing beam splitter. The light from the trapped microsphere with the unique polarization is then detected on one of the PSDs enabling force detection. To detect the force on the second trapped microsphere, the optical path of a 980 nm laser (IQ2C140, Power Technology) is then overlapped with the path of one of the three traps with identical polarization. Force detection is achieved by detecting the 980 nm light on a second PSD. Microsphere-to-microsphere distances are measured by template-matching on images created by LED on a complementary metal-oxide semiconductor (CMOS) camera (DCC1545M, Thorlabs). Fluorescence excitation is performed in an epi-, wide-field illumination by a 491 nm laser (Calypso 50, Cobolt), a 532 nm laser (Compass 215M-20, Coherent) and a 635 nm laser (IQ1C10(635-12B), Power Technology). The emitted fluorescence signal is detected after a band-pass filter and imaged on an EMCCD camera (Cascade 512 B, Princeton Instruments).

DNA, proteins and buffers. End-biotinylated bacteriophage λ DNA was connected to streptavidin-coated polystyrene beads to generate the DNA constructs as described previously[23]. For the experiments with DNA blunt ends, the restriction enzyme PmeI (NEB) was used to cleave the DNA and generate blunt ends. The streptavidin microspheres (4.5 μm diameter, Spherotech) were attached to the ends of the DNA using laminar flow[23].

All recombinant proteins have been expressed and purified from *Escherichia coli* cells. Biochemical analysis of the XRCC4 and XLF variants are described in Extended Data Figs 1–3 (expression vectors are available upon request) using previously established procedures[13,24]. Concentrations indicated throughout the text are the injected concentrations. Because a fraction of the protein is depleted from the solution as it sticks to the plastic and glass in the flow-system (u-Flux, LUMICKS B.V.), these are considered upper limits of the actual concentration present during measurements.

XRCC4 fluorescent labelling was performed with maleimide Alexa Fluor 555 at position 93 or position 218. XLF fluorescent labelling was performed with maleimide ATTO 647N at position 148 or with eGFP. For all protein samples used, protein activity was similar to wild-type activity. This was confirmed by biochemical DNA bridging assays (Extended Data Figs 1–3). Mass spectrometry analysis was performed in order to quantify the degree of labelling of the proteins (Extended Data Figs 1–3). In single-molecule assays, no difference was detected between the two different XRCC4 labelling positions (position 93 or 218).

Mass analyses were performed on a MALDI–TOF-TOF Bruker Ultraflex III spectrometer (Bruker Daltonics) controlled by the Flexcontrol 3.0 package (Build 51). This instrument was used at a maximum accelerating potential of 25 kV and was operated in linear mode and the m/z range from 20,000 to 80,000. The laser frequency was fixed to 100 Hz and approximately 1,000 shots by sample were cumulated. Five external standards (Protein Calibration Standard II, Bruker Daltonics) were used to calibrate each spectrum to a mass accuracy within 100 p.p.m. Peak picking was performed with Flexanalysis 3.0 software (Bruker) with an adapted analysis method. One microlitre of sample was mixed with 1 μl of a saturated solution of Sinapinic Acid in an acetonitrile/0.1%TFA mixture (30:70). Then 1 μl was spotted on the target, dried and analysed.

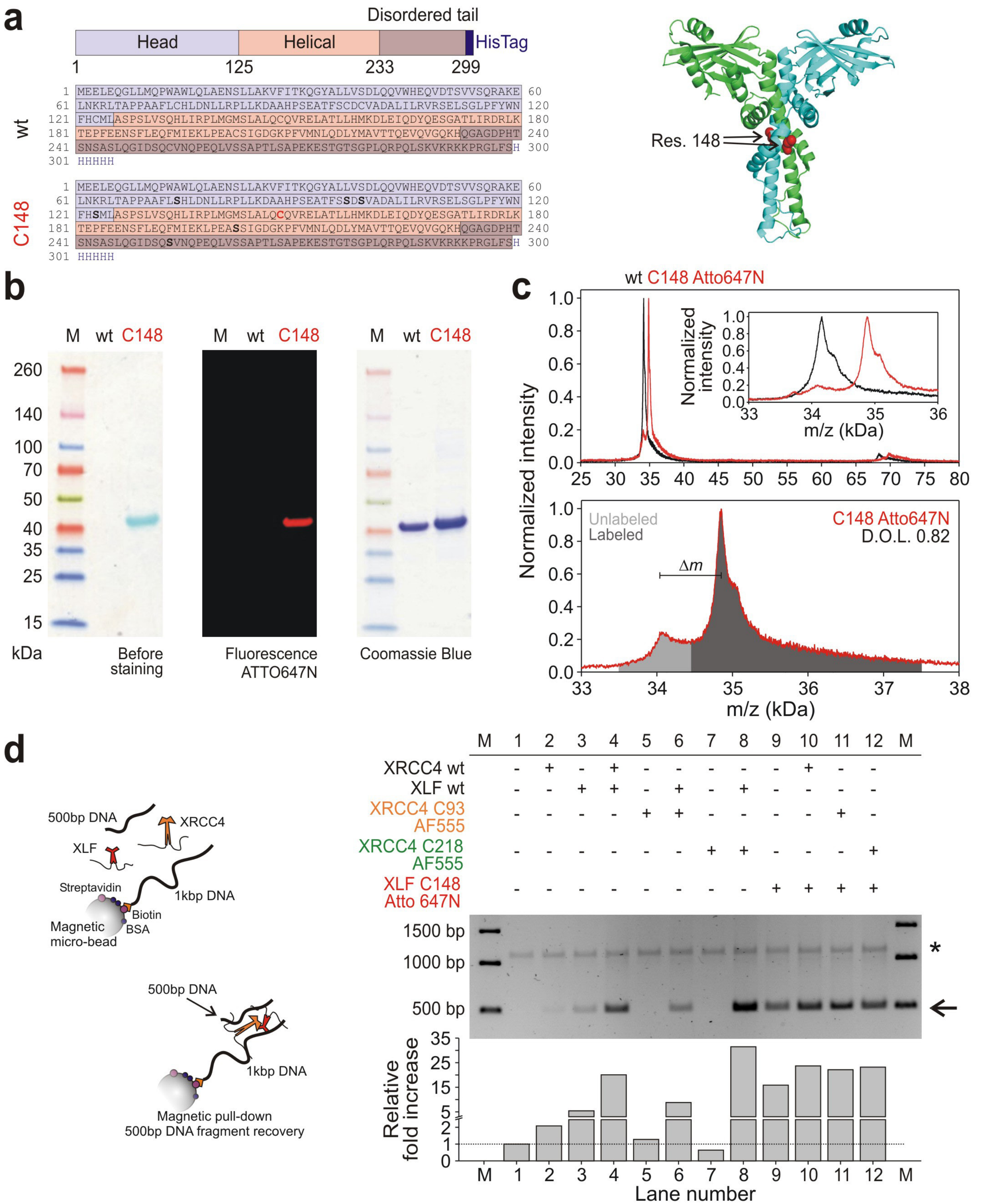
Protein–DNA interactions and force-stretching experiments were conducted in 20 mM Tris-HCl (pH 7.5), 1 mM DDT and, unless otherwise indicated, at 25 mM KCl.

The restriction enzyme SfoI used for the induction of DSBs was obtained commercially from NEB. Experiments were performed in 40 U ml^{-1} SfoI in CutSmart buffer supplied with the enzyme.

Protein binding. Binding of XRCC4 and XLF was quantified by quantifying the fluorescence. This was done by determining the average amount of photons per second and per watt of excitation power. All values were background corrected. Note that due to the difference in quantum and instrumental detection efficiency of eGFP and Alexa Fluor 555 the absolute amount of binding of XRCC4 (Fig. 2c) and XLF (Fig. 2g) cannot be compared directly.

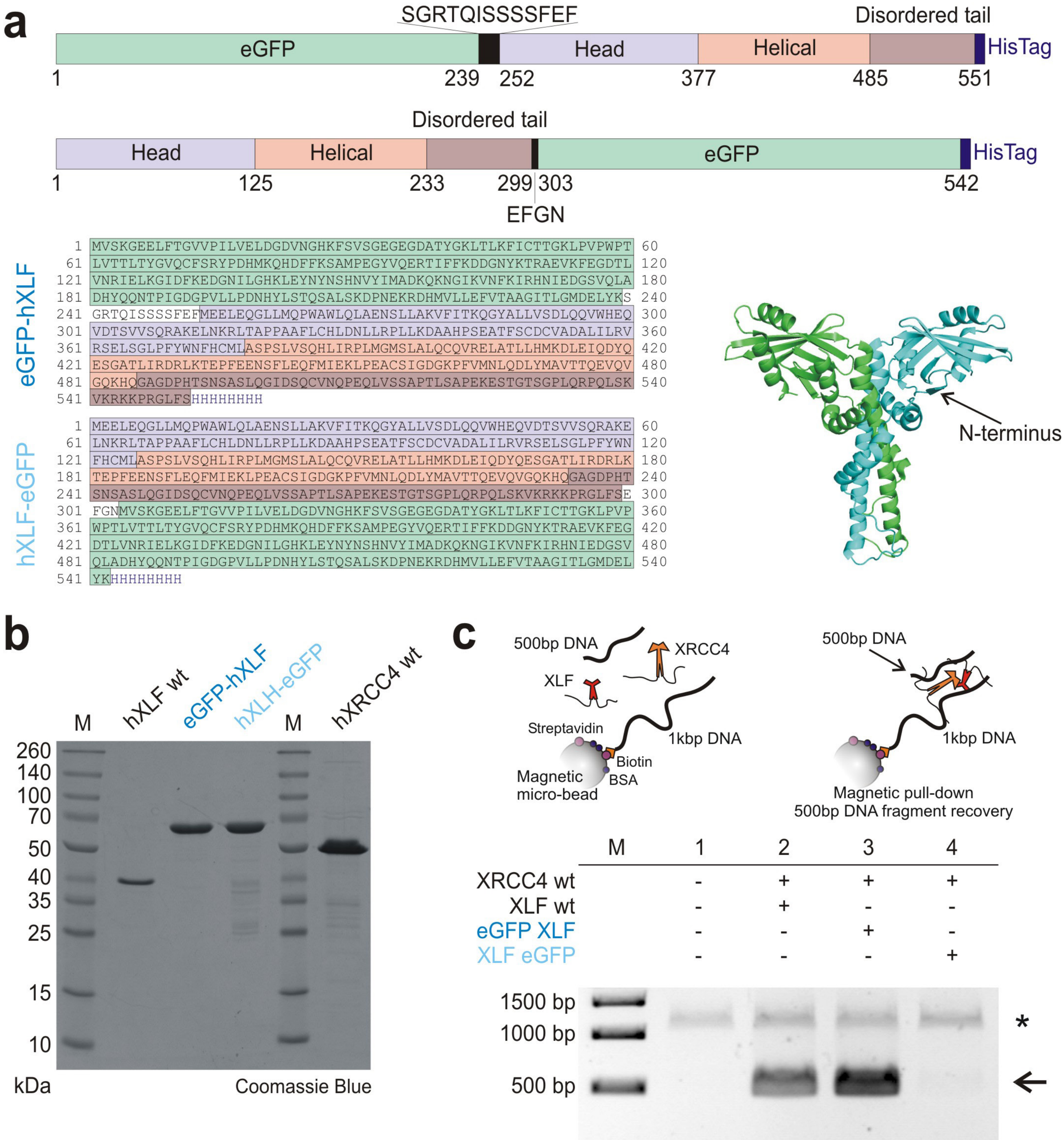
Protein diffusion. The one-dimensional diffusion of protein complexes along the DNA was quantified by tracking the mobile proteins for a large number of frames (typically more than 100) and calculating their diffusion coefficient (D) by using a mean square displacement analysis (MSD)[25] (Extended Data Fig. 6b). The diffusion coefficient obtained at low salt concentrations (25 mM KCl) was $0.73 \pm 0.09\,\mu m^2\,s^{-1}$ for XLF complexes, $0.68 \pm 0.09\,\mu m^2\,s^{-1}$ for XRCC4 complexes and $0.6 \pm 0.2\,\mu m^2\,s^{-1}$ for complexes containing both XRCC4 and XLF. These values are independent of DNA tension (Extended Data Fig. 6c) and complex size (Extended Data Fig. 6d), but increase with increasing salt concentration: at 160 mM KCl, the diffusion constant was $2.0 \pm 0.2\,\mu m^2\,s^{-1}$ for XRCC4 complexes.

22. Laurens, N. *et al.* Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat. Commun.* **3,** 1328 (2012).
23. Gross, P., Farge, G., Peterman, E. J. G. & Wuite, G. J. L. Combining optical tweezers, single-molecule fluorescence microscopy, and microfluidics for studies of DNA-protein interactions. *Methods Enzymol.* **475,** 427–453 (2010).
24. Andres, S. N., Modesti, M., Tsai, C. J., Chu, G. & Junop, M. S. Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Mol. Cell* **28,** 1093–1101 (2007).
25. Heller, I. *et al.* Mobility analysis of super-resolved proteins on optically stretched DNA: comparing imaging techniques and parameters. *ChemPhysChem* **15,** 727–733 (2014).

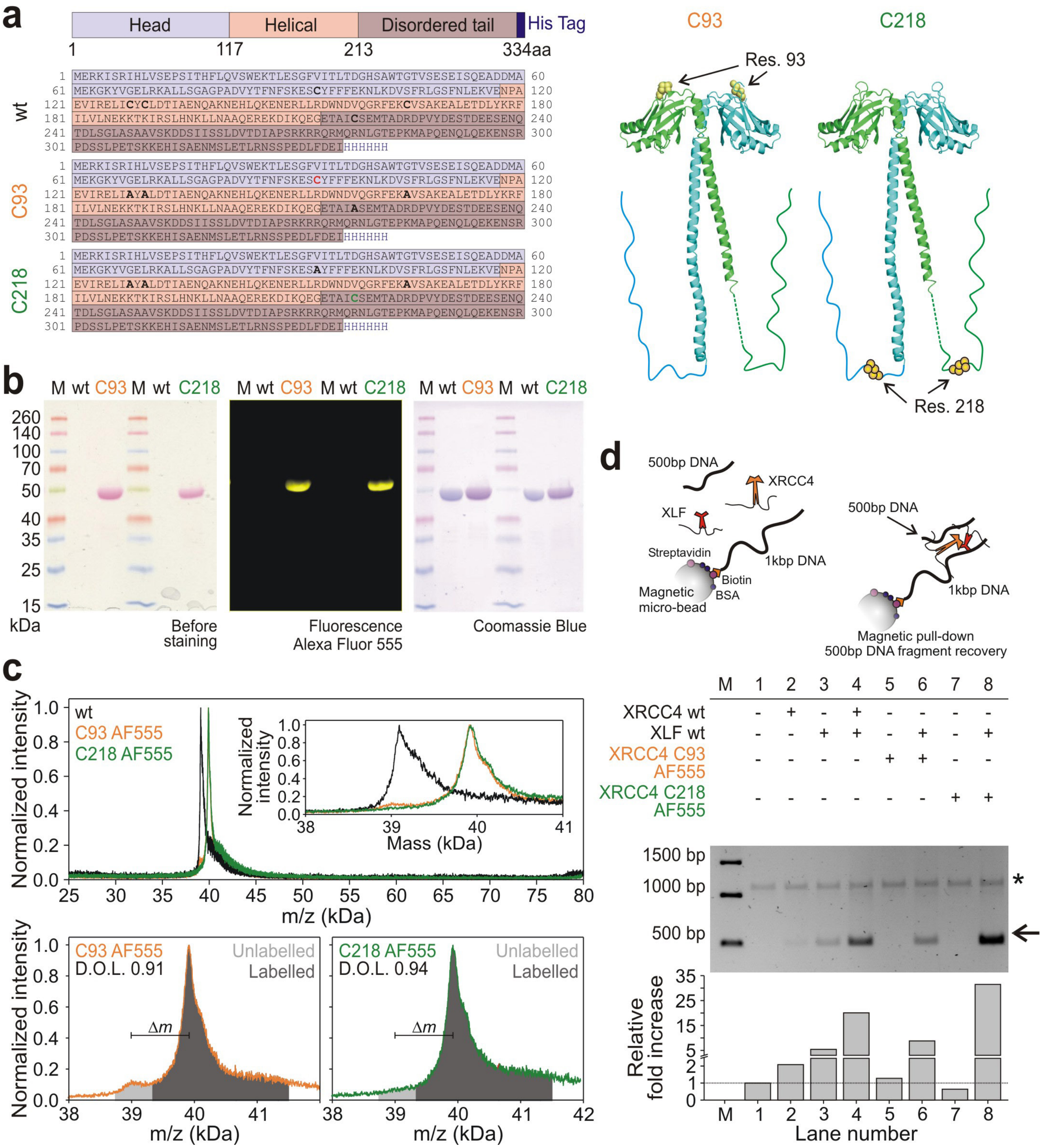


Extended Data Figure 1 | See next page for caption.

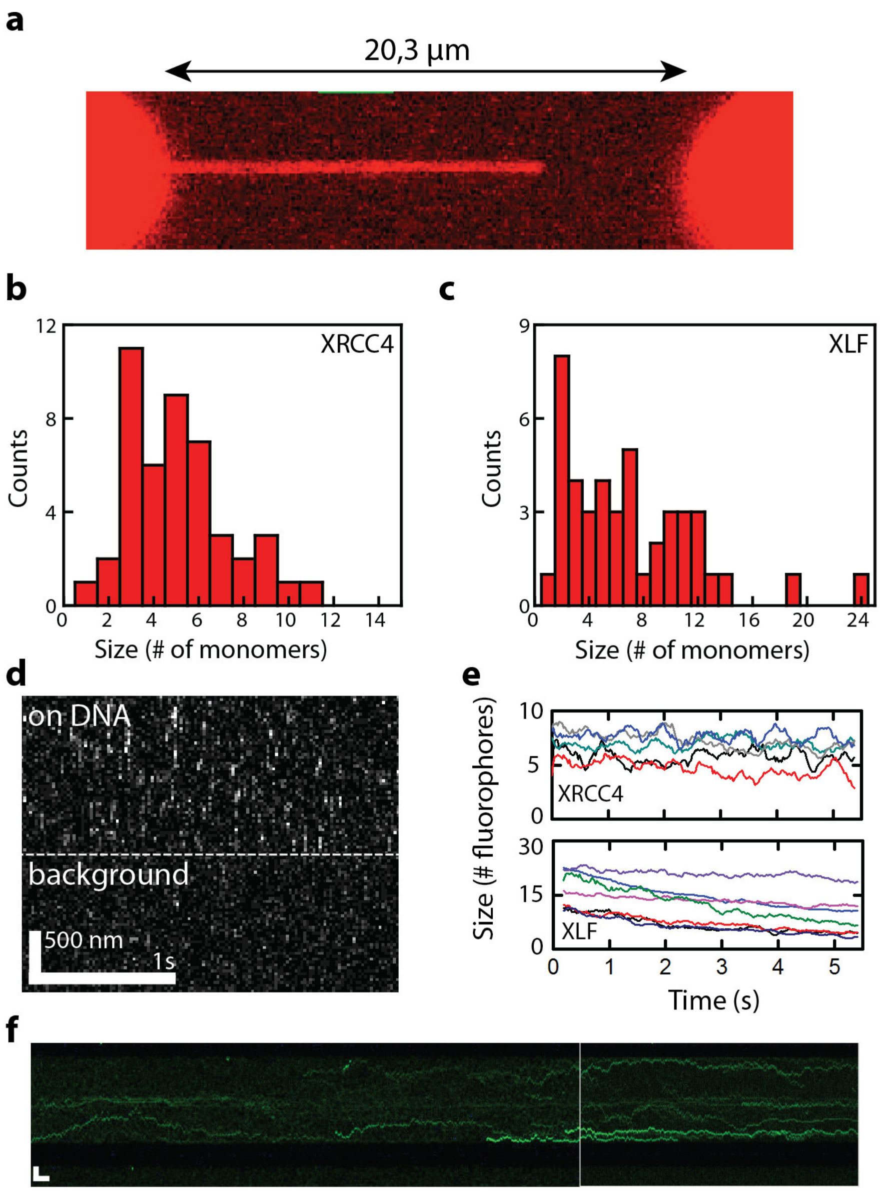
Extended Data Figure 1 | Generation of human XLF ATTO-647N-labelled fluorescent variant by the 'Cys light' method. a, Domain architecture of the XLF polypeptide and amino acid sequences of the wild-type and Cys148 variant obtained by site-directed mutagenesis. All Cys residues were changed to Ser except Cys148, leaving a single solvent accessible Cys residue. Three-dimensional model of the XLF dimer where the position–specific labelling sites are indicated by the black arrows. **b**, Denaturing and reducing polyacrylamide gel electrophoresis of XLF Cys148 variant after labelling with ATTO 647N maleimide. Left, bright-field image of the gel before staining. Centre, the emission of ATTO 647N. Right, is an image of the gel after staining with Coomassie. M, molecular mass markers. **c**, Mass spectrometry of the labelled full-length protein. Top, the spectra of XLF Cys148 (red) after labelling with ATTO 647N and compared to the wild-type unlabelled protein (black) giving $\Delta m = 790$ Da. Bottom, the calculation of the degree of labelling (D.O.L.), as the relative integrated intensity of the labelled protein (dark grey area) versus the unlabelled protein (light grey area) signals, giving 0.82 dye per monomer. **d**, DNA bridging activity of the XLF variants. Left, a scheme of the bridging assay in which an end-biotinylated 1,000-bp DNA fragment is coupled to streptavidin-coated magnetic beads. Protein-mediated DNA bridging is assessed by recovery of an unlabelled 500-bp DNA fragment. Right, image of an agarose gel stained with ethidium bromide to detect the bridged 500-bp DNA fragment (indicated by the black arrow) and the corresponding quantification. Asterisk indicates biotinylated 1,000-bp DNA fragment adsorbed non-specifically on the surface of the magnetic beads.



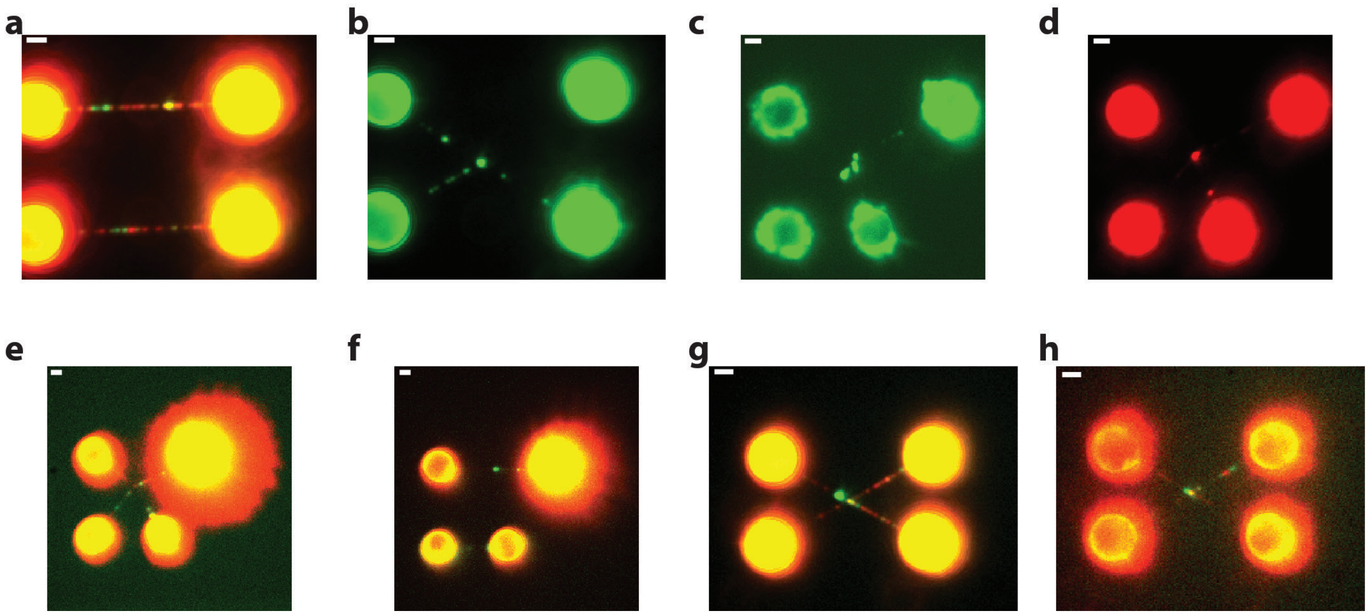
Extended Data Figure 2 | Generation of human XLF fluorescent variants by eGFP tagging. a, Domain architecture of the XLF polypeptide and amino acid sequences of the C- and N-terminal eGFP fusions. **b**, Denaturing and reducing polyacrylamide gel electrophoresis of the purified eGFP fusions. **c**, DNA bridging activity of the XLF variants. Top, a scheme of the bridging assay where an end-biotinylated 1,000-bp DNA fragment is coupled to streptavidin-coated magnetic beads. Protein-mediated DNA bridging is assessed by recovery of an unlabelled 500-bp DNA fragment. Bottom, an image of an agarose gel stained with ethidium bromide to detect the bridged 500-bp DNA fragment (indicated by the black arrow). Asterisk indicates biotinylated 1,000-bp DNA fragment adsorbed non-specifically on the surface of the magnetic beads.



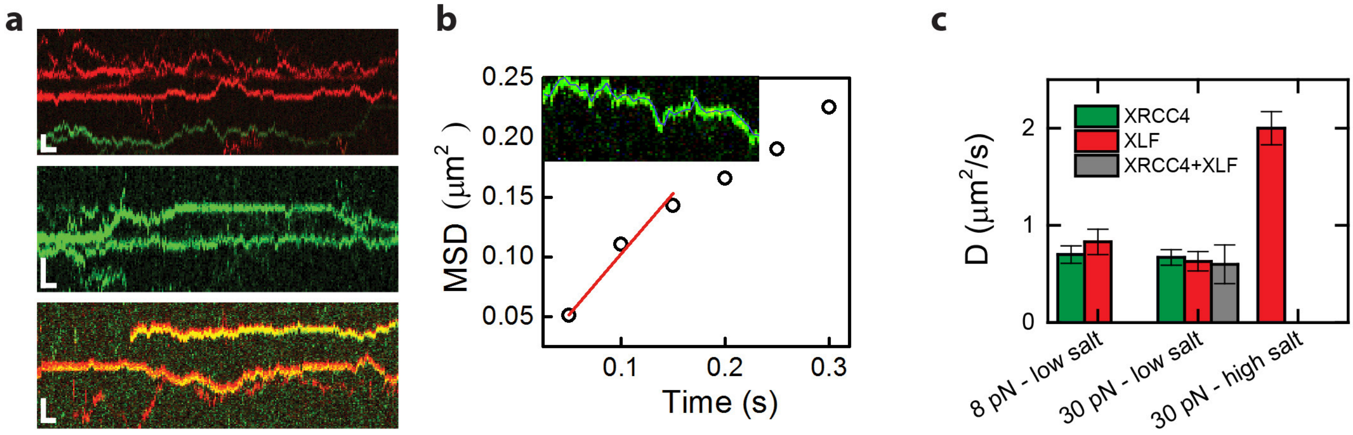
Extended Data Figure 3 | Generation of human XRCC4 fluorescent variants by the 'Cys light' method. a, Domain architecture of the XRCC4 polypeptide and amino acid sequences of the wild-type, C93 and C218 variants obtained by site-directed mutagenesis. All Cys residues were changed to Ala except, respectively C93 and C218, leaving for each variant a single solvent accessible Cys residue. Three-dimensional models of the XRCC4 dimer where the position-specific labelling sites are indicated by the black arrows. **b**, Denaturing and reducing polyacrylamide gel electrophoresis of XRCC4-C93 and -C218 variants after labelling with Alexa Fluor 555 maleimide. Left, bright-field image of the gel before staining. Centre, the emission of Alexa Fluor 555. Right, an image of the gel after staining with Coomassie. **c**, Mass spectrometry of the labelled full-length proteins. Top, the spectra of XRCC4-C93 (orange) and XRCC4-C218 (green) after labelling with Alexa Fluor 555 and compared to the wild-type unlabelled protein (black), giving $\Delta m = 830$ Da and $\Delta m = 835$ Da for labelled XRCC4-C93 and XRCC4-C218, respectively. Bottom, the calculation of the degree of labelling, as the relative integrated intensity of the labelled protein (dark grey area) versus the unlabelled protein (light grey area) signals, giving 0.91 and 0.94 dye per monomer for XRCC4-C93 and XRCC4-C218, respectively. **d**, DNA bridging activity of the XRCC4 variants. Top, a scheme of the bridging assay in which an end-biotinylated 1,000-bp DNA fragment is coupled to streptavidin-coated magnetic beads. Protein-mediated DNA bridging is assessed by recovery of an unlabelled 500-bp DNA fragment. Bottom, image of agarose gel stained with ethidium bromide to detect the bridged 500-bp DNA fragment (indicated by the black arrow) and the corresponding quantification (the image is from the same gel as in Extended Data Fig. 1d but only the lanes relevant to the analysis of fluorescently labelled XRCC4 are shown). Asterisk indicates biotinylated 1,000-bp DNA fragment adsorbed non-specifically on the surface of the magnetic beads.

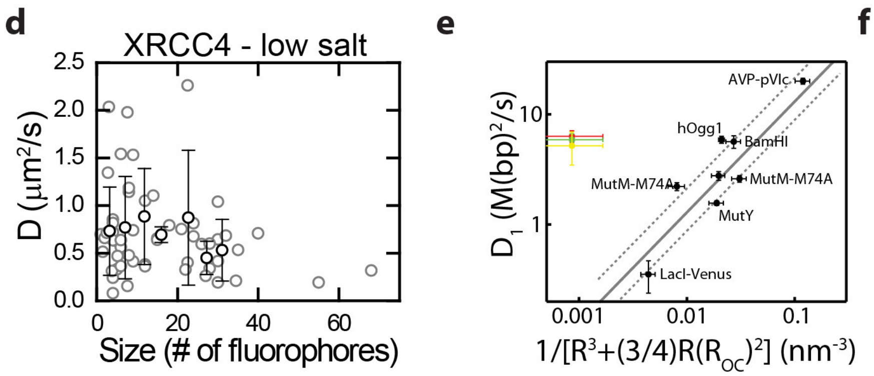


Extended Data Figure 4 | Quantification of XRCC4 and XLF binding to dsDNA. a, Fluorescence image of an overstretched dsDNA molecule in the presence of 100 nM eGFP–XLF. Under the given salt conditions and at this DNA extension, significant DNA melting and formation of ssDNA is expected. XLF shows a high affinity to bind to dsDNA (bright fluorescent signal), whereas it does not appear to bind to ssDNA (dark region). **b**, **c**, Size of the DNA-bound protein complexes for XRCC4 (**b**) and XLF (**c**) as determined from their fluorescence intensities (after incubation with 50 nM XRCC4 (**b**) or XLF (**c**)). **d**, Sections of a kymograph measured in the protein channel (25 nM eGFP–XLF). Short protein-binding events are visible as brief, local bursts of fluorescence. Events are shorter than the line scan time (10 ms). **e**, Typical intensity time traces of bound XRCC4 and XLF indicate that the complexes bind as a whole from solution and not monomer by monomer. The decrease of the fluorescence intensity in time is due to photobleaching of the fluorophores. **f**, Two successive kymographs (separated by white line). Two separate kymographs are recorded owing to technical limitations on the maximum recoding time of XRCC4–Alexa Fluor 555 binding measured using a very low excitation power to reduce the effect of photobleaching. Under these conditions, XRCC4 oligomers stay bound for long periods of time (in the order of several minutes). Scale bars, 5 s and 2 µm. Data are representative examples of 7 (**a**), 13 (**d**) and 5 (**f**) experiments.



Extended Data Figure 5 | Properties of protein bridges. a, **b**, Bridging also occurs in the absence of free protein in solution. Fluorescence images (XRCC4–Alexa Fluor 555 and XLF–ATTO 647N fluorescence in **a**, and XRCC4–Alexa Fluor 555 fluorescence in **b**) acquired in the absence of free protein in solution of two XRCC4–XLF-coated dsDNA molecules before (**a**) and after (**b**) wrapping and subsequent unwrapping shows that bridging does not require the presence of free protein in solution. Representative example out of 2 experiments. **c**, **d**, DNA–protein bridges also occur in the presence of either XRCC4 or XLF. Fluorescence images of bridges formed by wrapping and subsequent unwrapping in the presence of XRCC4–Alexa Fluor 555 (**c**) or eGFP–XLF (**d**). Representative example out of 10 (**c**) or 19 (**d**) experiments. **e**, **f**, In ~5% of the cases, rupture of protein bridges resulted into two intact dsDNA molecules bound with XRCC4–Alexa Fluor 555 and XLF–ATTO 647N. Fluorescence images before (**e**) and after (**f**) rupture of such a protein bridge. Representative example out of 9 experiments. **g**, **h**, Fluorescence images taken of an XRCC4–XLF bridge at $t = 0$ min (**g**) and $t = 95$ min (**h**), showing that the bridge and the protein complexes remain stably bound to the DNA segments. Green, XRCC4; red, XLF, yellow, colocalization. Scale bars, 2 μm.

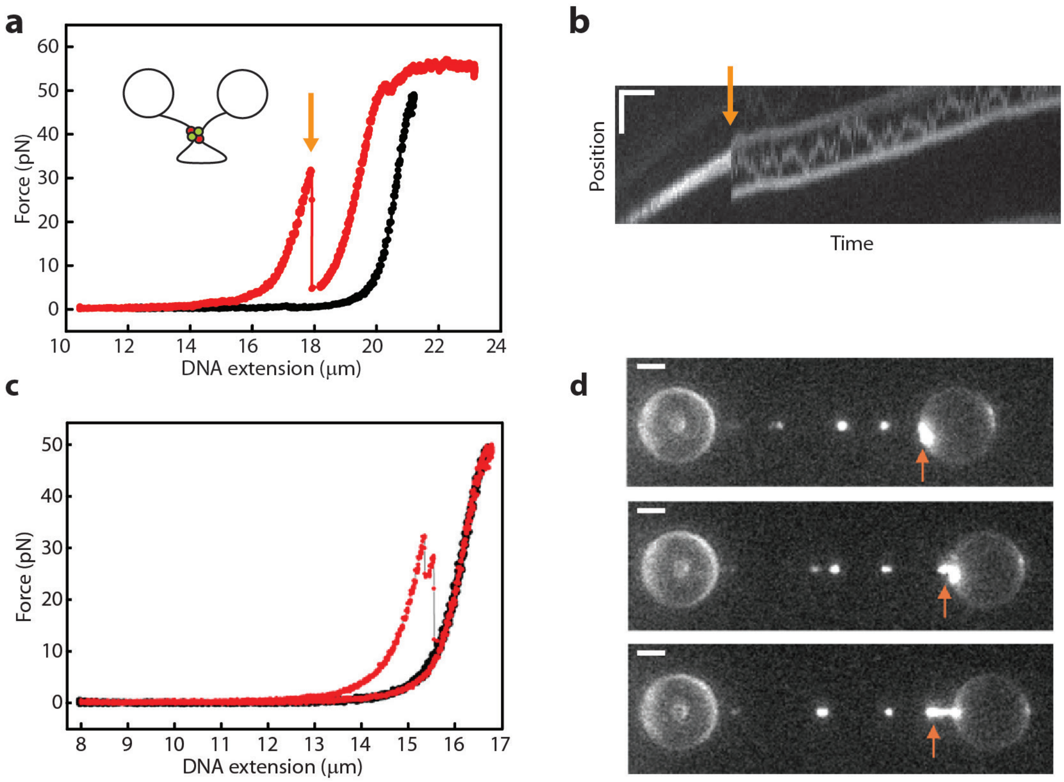




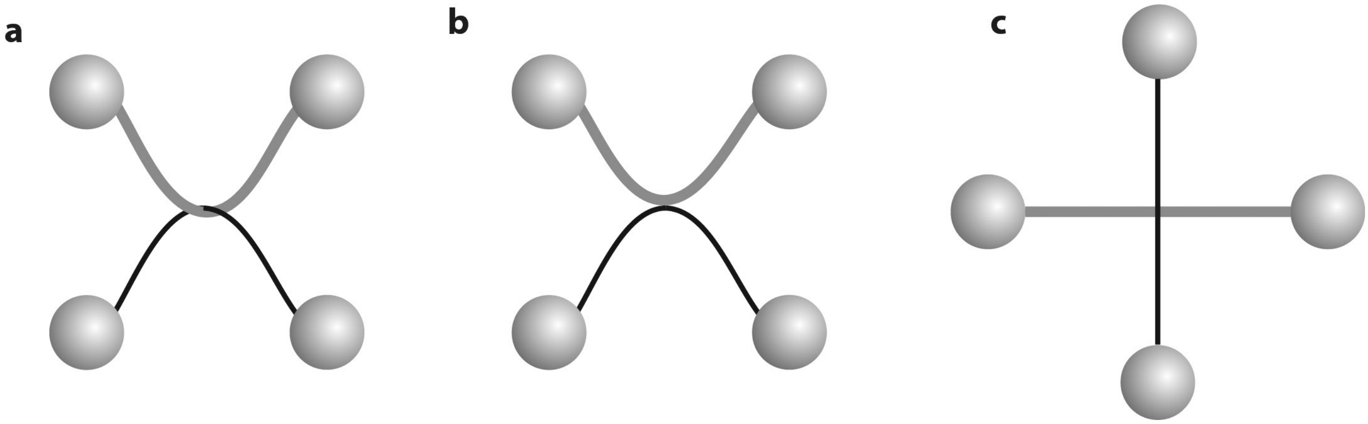
f

Protein	State	Average observed switching time (s)
XLF alone	Static	9±2
XLF alone	Diffusive	6±1
XRCC4 alone	Static	6±2
XRCC4 alone	Diffusive	7±2
XRCC4-XLF	Static	10±4
XRCC4-XLF	Diffusive	7±2

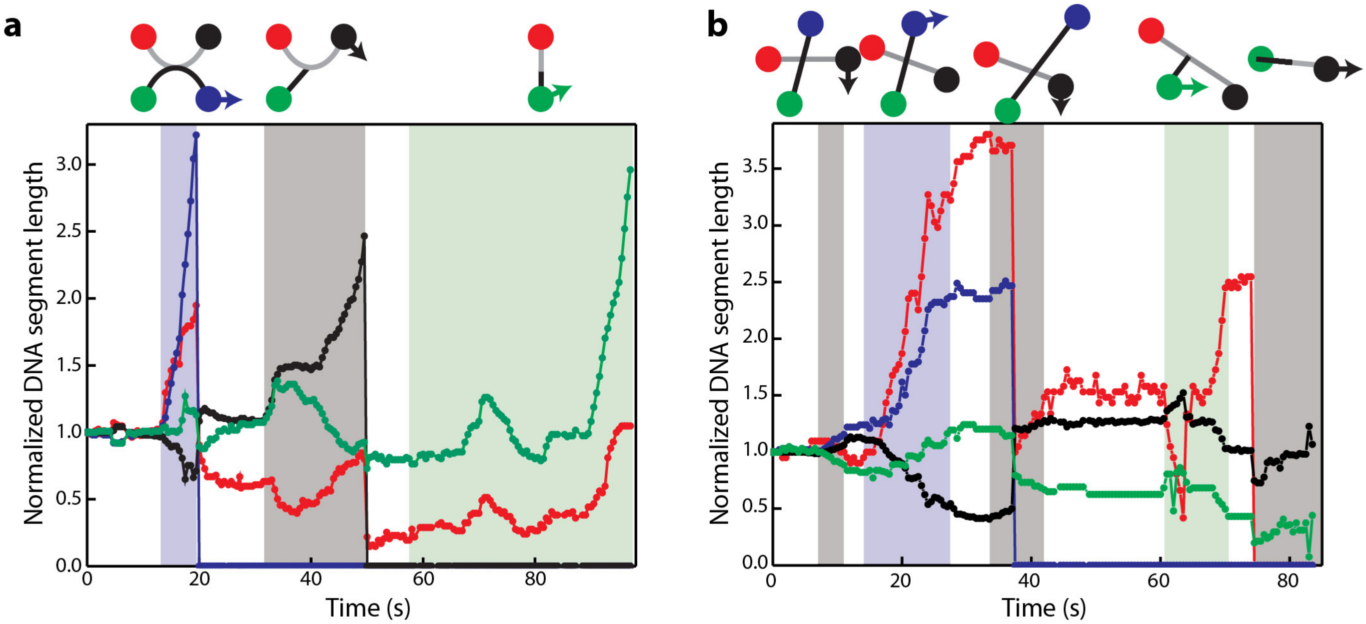
Extended Data Figure 6 | Quantification of diffusion behaviour of XRCC4 and XLF on dsDNA. a, Typical kymographs showing switching of XLF–ATTO 647N (red signal, top) XRCC4–Alexa Fluor 555 (green signal, middle) and XLF–XRCC4 (yellow signal, bottom) complexes between static and diffusive states. Scale bars, 1 s and 2 μm. **b**, Typical MSD curve of an individual XRCC4–Alexa Fluor 555 complex. Inset, kymograph of corresponding complex. Red line denotes linear fit to the first three data points; from the slope, a diffusion coefficient of 0.51 $\mu m^2 s^{-1}$ is determined. See Methods for details on MSD analysis. **c**, Diffusion coefficients of mobile XRCC4 and XLF complexes at different salt conditions (low salt: 25 mM KCl, high salt: 160 mM KCl) and DNA tensions. **d**, Diffusion coefficients of mobile XRCC4 protein complexes as a function of complex size. Grey circles, individual data points; black circles, average of 8 successive data points. **e**, Comparison of the observed diffusion coefficients of XRCC4 (green), XLF (red) and XRCC4–XLF complexes (yellow) to the expected diffusion coefficients based on the helical diffusion model[19] (solid black line). To calculate the quantity on the horizontal axis, a hydrodynamic radius of 7–22 nm was used, based on the inner and outer radii of the XRCC4–XLF filament as proposed previously[13]. **f**, Average observed dwell times of protein complexes before switching to a different mode. The analysis is performed on 121 events. The shortest dwell time that could be determined with certainty was 1 s. All error bars denote s.e.m.



Extended Data Figure 7 | Rupture events in force-extension curves can be caused by DNA–protein bridges or nonspecific sticking of protein-bound DNA to the trapped microspheres. a, **c**, Force-extension curves of dsDNA–XRCC4–XLF complexes (red) after incubation at low tension shows rupture events. Black data sets show force-extension curve of bare dsDNA. **b**, A single, continuous fluorescence kymograph corresponding to red curve in **a**. At the indicated time (orange arrow in **a** and **b**) a protein bridge suddenly ruptures into multiple smaller protein complexes. **d**, Fluorescence kymograph corresponding to red curve in **c** reveals that rupture event corresponds to the rupture of a non-specific interaction between a DNA-bound protein complex and the polystyrene microsphere. Scale bars, 1 s and 2 μm.

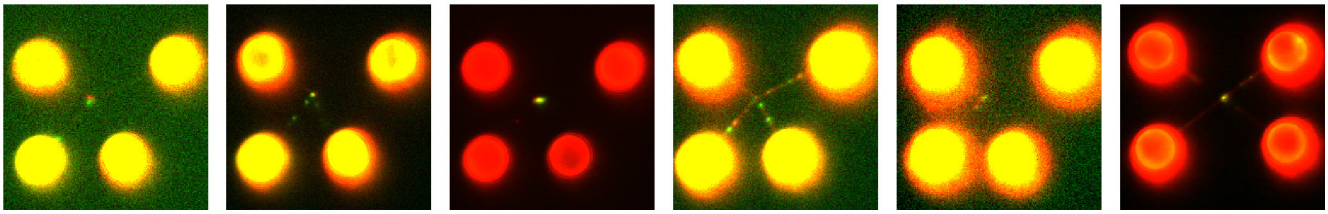


Extended Data Figure 8 | Schematic representation of dual-DNA configurations used in quadruple-trap experiments. a–c, Wrapped (**a**), unwrapped (**b**) and crossed (**c**) DNA configurations.



Extended Data Figure 9 | Quantification of DNA bridging by XRCC4–XLF. a, **b**, Normalized length of DNA segments over time during experiments such as shown in Figs 2 and 3. **a**, The static bridge in the experiment described in Fig. 2 and Supplementary Video 1. **b**, The mobile bridge in the experiment described in Fig. 3 and Supplementary Video 2.

a



Extended Data Figure 10 | Protein bridges always contain both XRCC4 and XLF. a, Typical examples of XRCC4–XLF bridges formed by incubating two wrapped DNA molecules in 200 nM eGFP–XLF and 200 nM XRCC4–Alexa Fluor 555 for 2 min. In all 152 bridges that were analysed, clear colocalization of the proteins at the junction is observed.

doi:10.1038/nature18647

Dynamics of ribosome scanning and recycling revealed by translation complex profiling

Stuart K. Archer[1,2]*, Nikolay E. Shirokikh[1,3]*, Traude H. Beilharz[4] & Thomas Preiss[1,5]

Regulation of messenger RNA translation is central to eukaryotic gene expression control[1]. Regulatory inputs are specified by the mRNA untranslated regions (UTRs) and often target translation initiation. Initiation involves binding of the 40S ribosomal small subunit (SSU) and associated eukaryotic initiation factors (eIFs) near the mRNA 5′ cap; the SSU then scans in the 3′ direction until it detects the start codon and is joined by the 60S ribosomal large subunit (LSU)[2–5] to form the 80S ribosome. Scanning and other dynamic aspects of the initiation model have remained as conjectures because methods to trap early intermediates were lacking. Here we uncover the dynamics of the complete translation cycle in live yeast cells using translation complex profile sequencing (TCP-seq), a method developed from the ribosome profiling[6] approach. We document scanning by observing SSU footprints along 5′ UTRs. Scanning SSU have 5′-extended footprints (up to ∼75 nucleotides), indicative of additional interactions with mRNA emerging from the exit channel, promoting forward movement. We visualized changes in initiation complex conformation as SSU footprints coalesced into three major sizes at start codons (19, 29 and 37 nucleotides). These share the same 5′ start site but differ at the 3′ end, reflecting successive changes at the entry channel from an open to a closed state following start codon recognition. We also observe SSU 'lingering' at stop codons after LSU departure. Our results underpin mechanistic models of translation initiation and termination, built on decades of biochemical and structural investigation, with direct genome-wide *in vivo* evidence. Our approach captures ribosomal complexes at all phases of translation and will aid in studying translation dynamics in diverse cellular contexts. Dysregulation of translation is common in disease and, for example, SSU scanning is a target of anti-cancer drug development[7]. TCP-seq will prove useful in discerning differences in mRNA-specific initiation in pathologies and their response to treatment.

Protein synthesis in eukaryotes is more complex than in prokaryotes, particularly at the highly regulated initiation step, in which eukaryote-specific mechanisms of ribosome attachment to mRNA and start codon recognition are used. Although multiple biochemical and structural studies support the scanning model of initiation, originally proposed by Marilyn Kozak almost forty years ago[8], there still is no direct *in vivo* evidence[2,5] and alternative mechanisms have been proposed[9]. Several issues remain difficult to resolve, such as whether scanning occurs in a cap-severed or cap-tethered manner; what mechanism facilitates directional movement of scanning SSU; what intermediates are adopted at particular stages; and what is the rate-limiting step of initiation *in vivo* (Extended Data Fig. 1 and Supplementary Discussion). Information on the position of translation complexes along mRNA is needed to address these questions. Ribosome profiling[6] maps elongating ribosomes along mRNA *in vivo,* however, it cannot record intermediates of translation initiation, termination and recycling.

To address this, we developed translation complex profile sequencing (TCP-seq; Fig. 1a and Supplementary Discussion). We crosslinked snap-chilled yeast (*Saccharomyces cerevisiae*) cells with formaldehyde, which efficiently fixes intermediates of ribosome function[10], to stall and attach any translation complex type to mRNA at their native positions. Next, polysomal translation complexes were isolated and digested with RNase I to generate protected mRNA fragments (footprints), as in ribosome profiling. However, ribosome and SSU fractions were then separated by sedimentation velocity before retrieval of footprints and high-throughput sequencing. Library inserts up to 250 nucleotides (nt) were selected to accommodate diverse footprint sizes and high-throughput sequencing reads (150 nt) were mapped to the yeast genome (Supplementary Table 1). Unfractionated polysomal mRNA

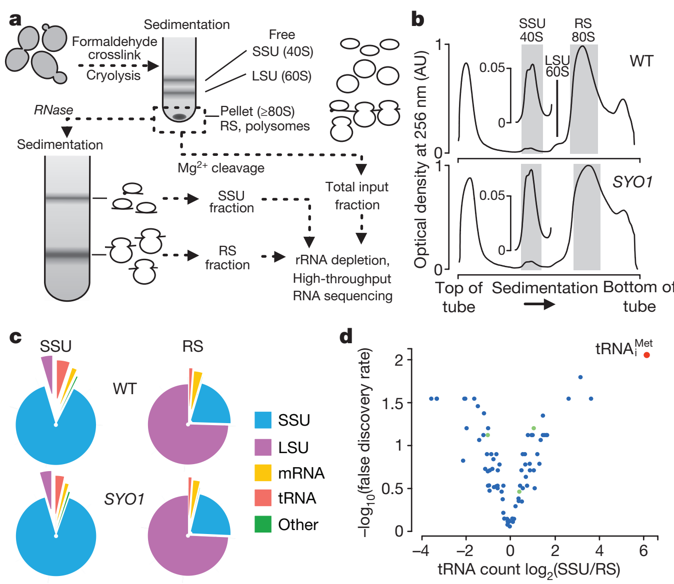


Figure 1 | Translation complex profile sequencing (TCP-seq) surveys all phases of mRNA translation. a, Outline of TCP-seq. See Methods and Supplementary Discussion for details. RS, ribosomes. **b**, Sedimentation profiles of ribosomes and ribosomal subunits after RNase cleavage of polysomes from wild-type (WT) and *SYO1*-TAP (*SYO1*) yeast strains. Shading indicates material used for high-throughput sequencing library preparation. Inserts show magnified (10×) SSU peaks. AU, arbitrary units. **c**, Distribution of mapped reads across RNA types from SSU and ribosome libraries (see Supplementary Table 1 for raw read numbers). Note that RNase digestion and broad library insert size selection during TCP-seq pose challenges for rRNA depletion (Supplementary Discussion). **d**, tRNA distribution between SSU and ribosome libraries. Averaged values from wild-type and *SYO1*-TAP libraries are shown. Initiator $tRNA_i^{Met}$ (red), elongator $tRNA_e^{Met}$ (green), and other tRNAs (blue).

[1]EMBL–Australia Collaborating Group, Department of Genome Sciences, The John Curtin School of Medical Research, The Australian National University, Canberra, Australian Capital Territory 2601, Australia. [2]Monash Bioinformatics Platform, Monash University, Melbourne, Victoria 3800, Australia. [3]Moscow Regional State Institute of Humanities and Social Studies, Kolomna 140410, Russia. [4]Development and Stem Cells Program, Monash Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Victoria 3800, Australia. [5]Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales 2010, Australia.
*These authors contributed equally to this work.

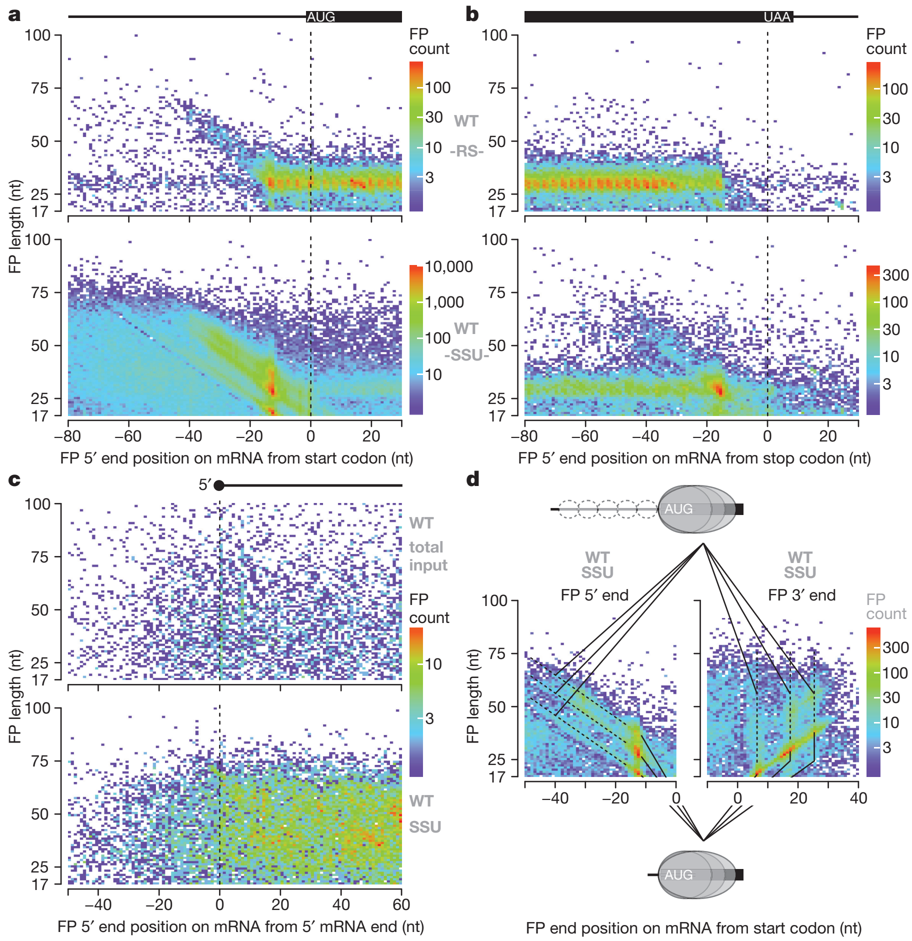


Figure 2 | Snapshots of SSU and ribosome footprint distribution reveal translation initiation by a cap-severed, SSU-led scanning mechanism and staged ribosome recycling after termination. Graphs depict cumulative mRNA data transcriptome-wide from wild-type cells (see Extended Data Fig. 5 for the *SYO1* strain results). Footprint (FP) 5′ ends are plotted versus footprint length; colour scale represents footprint count for each combination of footprint 5′ ends and footprint length (see Extended Data Fig. 6 for footprint 3′ end plots). **a**, Footprint distribution over the start codon region. All mRNAs are aligned by the first nucleotide (position 0) of the annotated ORF start codon. **b**, Footprint distribution over the stop codon region. As in **a**, but mRNAs are aligned by the first nucleotide (position 0) of the annotated ORF stop codon. **c**, Footprint distribution near mRNA cap structures. Only mRNAs with 5′ UTRs longer than 80 nt are included. **d**, Start codon focused plots of SSU footprint length versus footprint 5′ (left) or footprint 3′ (right) end positions on mRNAs with annotated 5′ UTRs longer than 100 nt. Dashed lines indicate three distinct classes of SSU footprints with shared footprint 3′ ends and variable footprint 5′ ends, as depicted in the schematic (more details and quantification are in Extended Data Fig. 6e, f). Here the ratio of footprint 5′ end density in a representative −70 nt to −50 nt window, to that on the start codon, is 1:21 (or 1:46 if footprints with 5′ ends between −16 nt and −10 nt are all included in the start codon tally).

was processed alongside as a control (total input). Steady-state polysomes mostly contain elongating ribosomes, with few initiating SSU. Accordingly, the sedimentation profiles featured a major ribosome-related peak and a minor SSU-related peak, the latter increased in mildly LSU-deficient *SYO1*-TAP (TAP, tandem-affinity purification) cells[11] (Fig. 1b). Libraries were sequenced from wild-type and *SYO1*-TAP cells. The latter replicated all qualitative findings, while displaying the expected quantitative differences due to LSU deficiency, as described below. Regarding the specificity of ribosome and SSU purification, SSU libraries contained mostly SSU ribosomal RNA, whereas ribosome libraries harboured LSU and SSU rRNA in a ∼3:1 ratio (Fig. 1c). Furthermore, initiator transfer RNA $\mathrm{tRNA_i^{Met}}$ was >60-fold enriched in SSU compared to ribosome libraries (Fig. 1d; Supplementary Table 1). Notably, ribosome footprint lengths, positions of footprint 5′ and 3′ ends relative to internal ribosome topology, and ribosome footprint distribution across open reading frames (ORFs) were all consistent with previous ribosomal profiling experiments[6,12] (Fig. 2a, b, top; Extended Data Figs 2 and 3a–c and Supplementary Discussion). No substantial sequencing bias was detected across the footprints (Extended Data Fig. 4). Altogether, this validates the TCP-seq approach.

Unlike ribosome footprints, SSU footprints were prominent in 5′ UTRs and enriched at start codons, with some accumulation also near stop codons (see below) and negligible presence in 3′ UTRs (Fig. 2a, b, bottom; Extended Data Figs 2a–d, f, 5a, b and 6a–d). SSU footprints within ORFs possessed characteristics indicating that they are due to disassembly of a small fraction (∼1%, as determined by sedimentation profiles, Fig. 1b) of elongating ribosomes during purification (Extended Data Fig. 3e, f and Supplementary Discussion), reflected also in a minor peak of LSU visible in Fig. 1b. The ratio of 5′ UTR to start codon footprint density was maximally 1:46 when aggregated across mRNAs having annotated 5′ UTRs long enough to unambiguously observe 5′ UTR footprints (see Fig. 2 legend and Supplementary Discussion). As expected, *SYO1*-TAP cells exhibited a similar SSU distribution to wild-type cells, but with stronger concentrations at start codons, consistent with delayed LSU joining in this strain (Extended Data Figs 2f and 5a–d). There was no increase in footprint density near the 5′ end of mRNAs (Fig. 2c; Extended Data Figs 2b, 7 and 8 for individual mRNAs); however, longer 5′ UTRs harboured more SSU on average (Extended Data Fig. 9a) without decreased SSU frequency at start codons (Extended Data Fig. 9b). There were no well-expressed mRNAs without SSU footprints in the 5′ UTR (Extended Data Fig. 9c).

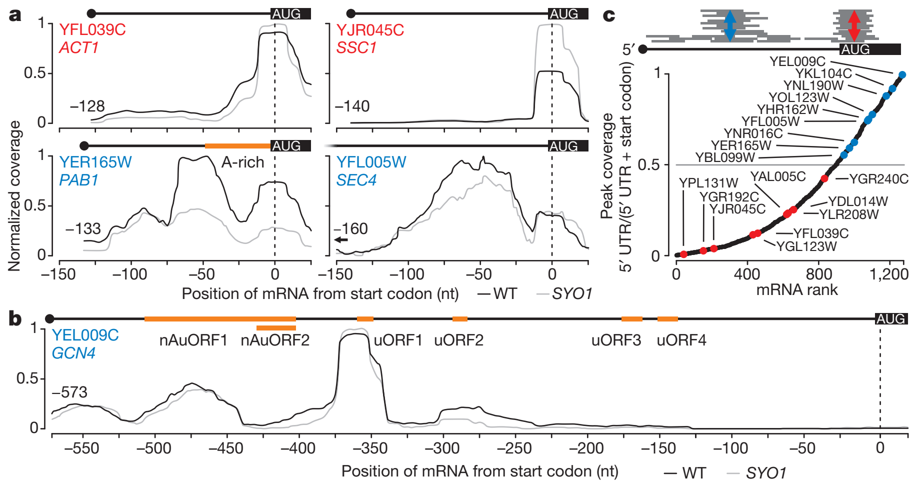


Figure 3 | TCP-seq can monitor modes and efficiency of translation initiation for individual mRNAs. a, **b**, Aggregated footprint coverage for wild-type (black) or *SYO1* (grey) cells versus 5′ UTR nucleotide position (aligned as in Fig. 2a). Names of individual mRNAs are indicated. Coverage is normalized to mRNA library size by the combined counts of 5′ UTR and start codon footprints for each mRNA and then scaled by the maximum value. Region of the *PAB1* mRNA sequence enriched for adenine nucleotides (**a**, bottom-left plot) and positions of the translated non-AUG (nAu) and canonical uORFs of the *GCN4* mRNA (**b**) are indicated by amber bars. For the complete footprint maps of these and other mRNAs, see Extended Data Figs 7 and 8, and the web interface for TCP-seq data display (http://bioapps.erc.monash.edu/TCP/). **c**, mRNAs were ranked by the ratio of peak SSU footprint coverage in the 5′ UTR, to sum of peak SSU footprint coverages in the 5′ UTR and at the start codon. Positions of individual mRNAs shown in Extended Data Figs 7 and 8 are indicated on the plot and colour-coded to reflect higher peaks on start codon (red) or in 5′ UTR (blue). See Supplementary Table 2 for a full list of ranked mRNAs.

These observations substantiate a model of cap-severed scanning (as opposed to cap-tethered scanning; Extended Data Fig. 1c) whereby SSU vacate 5′-end proximal regions of 5′ UTRs at a similar rate as they move along internal 5′ UTR regions, and multiple SSU can simultaneously scan a 5′ UTR.

Remarkably, 5′ UTR SSU footprint sizes ranged evenly from ~19 to ~75 nt (Fig. 2 and Extended Data Figs 2e and 6e), well beyond the width of SSU alone and reminiscent of *in vitro* observations in which edeine-enforced start codon bypass resulted in broader mRNA protection[13]. To assess where SSU reside within such extended scanning complexes, we inspected footprint patterns on approach to the start codon. There, complexes were uniformly arrested with their leading edges at one of three major positions (Fig. 2d, right; Extended Data Fig. 6e), consistent with the expected entry channel configurations of initiating SSU (see later for analysis of footprint 3′ patterns at start codons). Concomitantly, footprint 5′ density for extended complexes began to decline from ~40 nt upstream of start codons (Fig. 2d, left). This pattern indicates that the extended length of scanning SSU footprints is due to additional mRNA contacts upstream, but not downstream of SSU (Fig. 2d and Extended Data Fig. 6f), consistent with an earlier hypothesis that mRNA-binding by eIF4A, eIF4B/H and eIF4F prevents SSU from back-sliding[14,15] (Supplementary Discussion; Extended Data Figs 1d and 10). Intriguingly, we also noted a minor presence of the 5′-extended footprints among start codon-associated ribosomes, suggesting some carryover of the scanning eIFs into early elongation as postulated based on observations with reinitiation after translation of short ORFs[16] (Fig. 2a and Extended Data Figs 5a and 6a, c). Collectively, these observations favour an SSU-led, or 'pushed', model of scanning, whereby additional interactions preferentially occur on mRNA behind the SSU, enforcing movement in the 5′→3′ direction, as opposed to a 'pulled' SSU model, in which such contacts would occur in front of the SSU (Extended Data Figs 1d and 10a).

SSU footprints were further analysed for individual mRNAs with sufficient read depth. Multiple mRNAs with seemingly unstructured[17] and relatively short 5′ UTRs featured a smooth distribution of scanning SSU throughout their 5′ UTRs (including *ACT1* and *SSC1* in Fig. 3a; Extended Data Fig. 7), consistent with a linear, constant-rate scanning motion (Extended Data Fig. 1b); others displayed evidence of scanning block (Extended Data Fig. 8; Supplementary Discussion). This was most common with longer 5′ UTRs and often associated with short upstream ORFs (uORFs) at which the SSU accumulated (for example, *GCN4* in Fig. 3b). However, in many cases 5′ UTR SSU clustering could not be attributed to uORFs or apparent RNA structure[17], and these probably represent initiation control governed by specific factor binding (as can be expected for the A-rich sequence of *PAB1* mRNA[18]; Fig. 3a; Extended Data Fig. 8b and Supplementary Discussion) or the nucleotide context of the 5′ UTR itself. Overall, the full range of 5′ UTR to start codon ratios of peak coverage was observed (Fig. 3c), with some over-representation on the lower band due to the prevalence of 5′ UTRs that are too short for unambiguous detection of 5′ UTR SSU footprints. This indicates prevalent use of 5′ UTR features to 'tune' scanning, and thus initiation efficiency, across the yeast transcriptome. The set of mRNAs ranking highly in the ratio of 5′ UTR to start codon peak coverage is rich in both known examples of, and new candidates for, regulatory 5′ UTR elements (for example, *SEC4* mRNA in Fig. 3a; see also Supplementary Discussion, Supplementary Table 2, and the web interface for browsing and analysing the TCP-seq data at http://bioapps.erc.monash.edu/TCP/). Thus, TCP-seq has unique utility for surveying and dissecting initiation control mechanisms operating on individual mRNAs.

Start-codon-associated SSU footprints exhibited three main lengths (19, 29 and 37 nt; Fig. 2a bottom; Extended Data Figs 2e, 5a, 6a, c and 10a), mainly due to diversity in mRNA protection at the entry to the mRNA channel, with mostly uniform protection at the exit side (footprint 5′ at −12 nt, footprint 3′ at +6, +16 and +24 nt; Fig. 4a and Extended Data Fig. 5c). Recent structural studies can help interpret this pattern (Extended Data Fig. 10), with the caveat that *in vivo* initiation intermediates might not completely match the composition and/or structural configuration of *in vitro* assembled complexes. SSU footprints spanning from −12 nt to +6 nt (19 nt in length) conceivably derive from paused SSU still in open, scanning-competent configuration[19]. SSU footprints extending to +16 nt would represent SSU in the closed state following AUG recognition, protecting 29 nt of mRNA. These SSU are expected to feature disengaged eIF1 and the N terminus of eIF1A structured in the A-site[20,21] and an entry channel that is tightly closed on mRNA[19], with the eIF1A carboxy terminus bound to eIF5 (refs 20, 22). Finally, footprints ending at +24 nt could reflect a third SSU state just before LSU joining, where eIF5B:GTP attaches

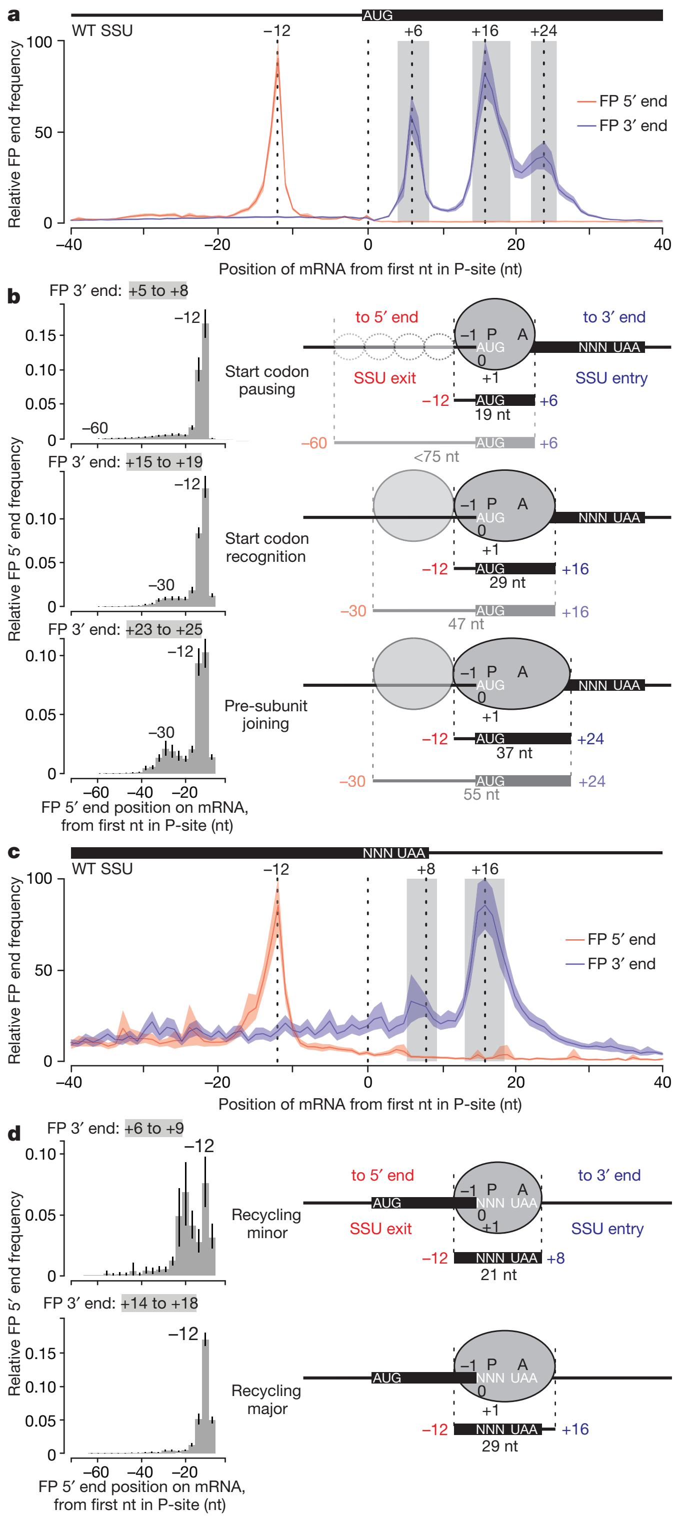


Figure 4 | Changes in SSU footprint size reveal dynamic rearrangements at start and stop codons. a, Aggregated profiles of SSU footprint 5′ end (red) and footprint 3′ end (blue) frequencies are plotted against mRNA position, and aligned by the first nucleotide in the P-site of start codon-arrested SSU (first nucleotide of the start codon) set at position 0. Footprint frequencies are scaled by the maximum values and peak frequency positions indicated. The error regions shown as shading of the same colour indicate the 95th percentiles of 1,000 iterations of random gene-wise resampling across all 5,859 represented genes (see Methods for details). Data shown are from all annotated mRNAs in wild-type cells; see Extended Data Fig. 5c, d for results with *SYO1*. **b**, Left, relative frequencies of the footprint 5′ ends corresponding to the +6 (top), +16 (middle) and +24 nt (bottom) major footprint 3′ end types plotted against position on mRNA, aligned as in **a**. Footprint 3′ end selection windows are indicated. Only mRNAs with annotated 5′ UTRs longer than 50 nt are included. Error bars depicted in black are calculated as in **a**. Right, schematics summarizing footprint 5′ and 3′ ends for the major SSU footprint types observed in the corresponding left panels. The proposed cause for the different 5′-extended footprint protection is also indicated (see main text and Extended Data Fig. 10 for details). **c**, Same as **a**, except the alignment is by first nucleotide in the P-site of terminating SSU (first nucleotide of the last sense codon). **d**, Left, same as **b**, but for the +8 (top) and +16 nt (bottom) footprint 3′ clusters of terminating SSU. These two footprint 3′ end peaks were most apparent in the footprint subset having footprint 5′ end localization characteristic of terminating complexes (Extended Data Figs 3d and 5e). Right, same as **b**, but for the stop-codon-associated SSU (UAA represents any of the three annotated stop codons).

at the entry of the mRNA channel, further extending protection[23,24]. Additional support for placement of SSU footprints in such a temporal sequence was derived from grouping them based on their 3′ ends and inspecting extended 5′ ends. Footprints ending around +6 nt included a minor population with footprint 5′ ends at up to −60 nt, suggesting that additional interactions characteristic of scanning are still partly in place (Fig. 4b, top). Footprints ending at +16 and +24 nt increasingly developed a second population with footprint 5′ ends around −30 nt, possibly the result of additional scanning SSU 'queuing' and further extending protection upstream by ~19 nt, due to cap recruitment and scanning outpacing start codon clearance (Fig. 4b, middle and bottom), although other explanations are also plausible. The shape of the SSU-related sedimentation peak (Fig. 1b) is also consistent with such complex heterogeneity. SSU footprints with all three footprint 3′ end positions are discernible on multiple individual mRNAs (Fig. 3a, b and Extended Data Figs 7, 8 and 9d–f), albeit in varying proportions probably indicative of an influence of local mRNA context on start codon recognition. These analyses serve to underpin models derived from biochemical and structural data, regarding the temporal sequence and molecular nature of rearrangements within SSU complexes at start codons, with direct *in vivo* evidence.

Finally, enrichment of SSU footprints of two main lengths (29 and 21 nt) is also observed on the last sense codon of annotated ORFs (Fig. 2b bottom; Extended Data Figs 2a, c, e, 5b, d and 6b, d, bottom panels). They share one footprint 5′ end at −12 nt, with footprint 3′ ends at +8 nt (minor) and +16 nt (major) (Fig. 4c; positions given relative to the first nucleotide in the P-site codon; see Extended Data Figs 3d and 5e), which is consistent with the increased presence of 3′-receded ribosome footprints (Extended Data Figs 3a, b and 5b, d). These shorter SSU and ribosome footprints conceivably reflect ribosome recycling intermediates with a more 'open' entry channel configuration, potentially resulting from ABCE1 activity at the entry site[25] (Extended Data Fig. 10b). A small fraction of terminating SSU in wild-type cells again feature footprints with 5′-extended protection (Fig. 2b). It is tempting to speculate that this reflects interactions of ABCE1 with eIF2, eIF5 (ref. 26) and eIF3 (ref. 27), and eRF3 with Pab1[28], potentially conferring these scanning factors on the expended SSU early during recycling to expedite its transfer for re-initiation *in cis* through the closed-loop mRNA structure formed by 5′ cap–eIF4E–eIF4G–Pab1–3′ poly(A) interactions. As a major conclusion, these findings provide direct evidence of subunit splitting before mRNA dissociation[29] for at least a considerable proportion of termination events, rather than complete ribosome dissociation[30] from mRNA (Extended Data Figs 1e and 10).

TCP-seq generates *in vivo* evidence that can be used to address general and mRNA-specific aspects of translational control, particularly by 5′ UTR elements. Several entirely new lines of research are now within reach. For example, the combination of TCP-seq with existing collections of yeast initiation factor mutant strains will refine roles of individual factors and substantiate the molecular nature of translation complex intermediates. Future studies should apply TCP-seq to

monitor translation under different growth conditions and in other eukaryotic and prokaryotic organisms. Given the importance of translational control to human disease, including cancer[7], TCP-seq could be used to aid in characterizing pathologic changes in translation and in the development of therapeutic approaches.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 December 2015; accepted 14 June 2016.
Published online 20 July 2016.

1. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
2. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83,** 779–812 (2014).
3. Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Rev. Mol. Cell Biol.* **11,** 113–127 (2010).
4. Hershey, J. W., Sonenberg, N. & Mathews, M. B. Principles of translational control: an overview. *Cold Spring Harb. Perspect. Biol*. **4,** a011528 (2012).
5. Merrick, W. C. Eukaryotic protein synthesis: still a mystery. *J. Biol. Chem.* **285,** 21197–21201 (2010).
6. Ingolia, N. T. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Rev. Genet.* **15,** 205–213 (2014).
7. Bhat, M. *et al.* Targeting the translation machinery in cancer. *Nature Rev. Drug Discov*. **14,** 261–278 (2015).
8. Kozak, M. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15,** 1109–1123 (1978).
9. Paek, K. Y. *et al.* Translation initiation mediated by RNA looping. *Proc. Natl. Acad. Sci. USA* **112,** 1041–1046 (2015).
10. Valásek, L., Szamecz, B., Hinnebusch, A. G. & Nielsen, K. H. *In vivo* stabilization of preinitiation complexes by formaldehyde cross-linking. *Methods Enzymol.* **429,** 163–183 (2007).
11. Li, Z. *et al.* Rational extension of the ribosome biogenesis pathway using network-guided genetics. *PLoS Biol.* **7,** e1000213 (2009).
12. Lareau, L. F., Hite, D. H., Hogan, G. J. & Brown, P. O. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* **3,** e01257 (2014).
13. Kozak, M. & Shatkin, A. J. Migration of 40S ribosomal subunits on messenger RNA in the presence of edeine. *J. Biol. Chem.* **253,** 6568–6577 (1978).
14. Siridechadilok, B., Fraser, C. S., Hall, R. J., Doudna, J. A. & Nogales, E. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310,** 1513–1515 (2005).
15. Spirin, A. S. How does a scanning ribosomal particle move along the 5′-untranslated region of eukaryotic mRNA? Brownian Ratchet model. *Biochemistry* **48,** 10688–10692 (2009).
16. Pöyry, T. A., Kaminski, A. & Jackson, R. J. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev.* **18,** 62–75 (2004).
17. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467,** 103–107 (2010).
18. Gilbert, W. V., Zhou, K., Butler, T. K. & Doudna, J. A. Cap-independent translation is required for starvation-induced differentiation in yeast. *Science* **317,** 1224–1227 (2007).
19. Llácer, J. L. *et al.* Conformational differences between open and closed states of the eukaryotic translation initiation complex. *Mol. Cell* **59,** 399–412 (2015).
20. Saini, A. K. *et al.* Eukaryotic translation initiation factor eIF5 promotes the accuracy of start codon recognition by regulating Pi release and conformational transitions of the preinitiation complex. *Nucleic Acids Res.* **42,** 9623–9640 (2014).
21. Lomakin, I. B. & Steitz, T. A. The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature* **500,** 307–311 (2013).
22. Luna, R. E. *et al.* The C-terminal domain of eukaryotic initiation factor 5 promotes start codon recognition by its dynamic interplay with eIF1 and eIF2β. *Cell Rep.* **1,** 689–702 (2012).
23. Yamamoto, H. *et al.* Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. *Nature Struct. Mol. Biol*. **21,** 721–727 (2014).
24. Fernández, I. S. *et al.* Molecular architecture of a eukaryotic translational initiation complex. *Science* **342,** 1240585 (2013).
25. Becker, T. *et al.* Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* **482,** 501–506 (2012).
26. Chen, Z. Q. *et al.* The essential vertebrate ABCE1 protein interacts with eukaryotic initiation factors. *J. Biol. Chem.* **281,** 7452–7457 (2006).
27. Khoshnevis, S. *et al.* The iron-sulphur protein RNase L inhibitor functions in translation termination. *EMBO Rep.* **11,** 214–219 (2010).
28. Roque, S. *et al.* Interaction between the poly(A)-binding protein Pab1 and the eukaryotic release factor eRF3 regulates translation termination but not mRNA decay in *Saccharomyces cerevisiae. RNA* **21,** 124–134 (2015).
29. Jackson, R. J., Hellen, C. U. & Pestova, T. V. Termination and post-termination events in eukaryotic translation. *Adv. Protein Chem. Struct. Biol.* **86,** 45–93 (2012).
30. Kurata, S. *et al.* Possible steps of complete disassembly of post-termination complex by yeast eEF3 deduced from inhibition by translocation inhibitors. *Nucleic Acids Res.* **41,** 264–276 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by an ARC Discovery Grant (DP1300101928) and an NHMRC Senior Research Fellowship (514904) awarded to T.P. N.E.S. was supported by a Go8 European Fellowship. We are grateful to A. G. Hinnebusch, C. G. Proud and R. D. Hannan for discussions and suggestions for this work. We acknowledge technical support from the Australian Cancer Research Foundation Biomolecular Resource Facility (JCSMR, ANU), D. Powell and S. Androulakis at the Monash Bioinformatics Platform.

Author Contributions S.K.A., T.H.B. and T.P. designed the research, S.K.A. and N.E.S. performed the experiments, S.K.A, N.E.S., T.H.B. and T.P. analysed the data, discussed the results and wrote the paper.

Author Information The sequences of reads with poly(A) tracts were deposited to the Sequence Read Archive, under accession code SRP074093. An online interface for browsing and analysing the TCP-seq data is available at (http://bioapps.erc.monash.edu/TCP/) along with (http://dx.doi.org/10.6084/m9.figshare.3206725). The underlying mapped dataset is also available (http://dx.doi.org/10.6084/m9.figshare.3206698). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.P. (thomas.preiss@anu.edu.au).

Reviewer Information *Nature* thanks E. Alkalaeva, P. Baranov and Y. Mechulam for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines. The wild-type yeast cell line was BY4741. The *SYO1*-TAP (YDL063C–TAP) mutant cell line was a BY4741 yeast strain containing a C-terminal insertion of the modified TAP-tag in YDL063C[31]. *SYO1*-TAP is known to exhibit cytoplasmic LSU shortage and decreased initiation efficiency due to longer LSU waiting times, also manifested as polysomal half-mer phenotype in sucrose gradient sedimentation profiles[11], similar to the *SYO1*-deficient cell lines[32].

Cell crosslinking and lysis. Wild-type and *SYO1*-TAP cells were each grown to an optical density at 600 nm (OD_{600} or A_{600nm}) of 0.7–0.8 in 1 litre of YPAD media and snap-chilled by mixing with 25% w/v of crushed ice. 0.1 volumes of freshly depolymerized 30% w/v paraformaldehyde was immediately added with intensive mixing and the mixture left for 10 min on ice. The fixed cells were pelleted by centrifugation at 4 °C, 5,000*g* for 5 min and resuspended in 40 ml of buffer A (20 mM HEPES-KOH pH 7.4 at 25 °C, 100 mM KCl and 2 mM $MgCl_2$) supplemented with 0.25 M glycine after 20 min of total fixation time. The cells were washed twice by sequential resuspension in 40 ml of buffer A and pelleting as described above. The washed cell pellet (1.1 g wet cell mass) was aspirated and resuspended in 40% v/w of buffer A supplemented with 0.5 mM DTT, 1 U μl^{-1} RNaseOUT RNase inhibitor (Life Technologies) and 1× Complete EDTA-free Mini Protease Inhibitor (Roche). The cell suspension was then flash-frozen by dripping into liquid nitrogen. A 10 ml stainless steel grinding jar (Retsch) was pre-cooled in liquid nitrogen and filled with ~2 g of the frozen cell suspension droplets and two 12 mm stainless steel grinding balls (Retsch). The sealed grinding jar was shaken at 27 Hz for one minute in the Mixer Mill MM 400 (Retsch), re-cooled in liquid nitrogen, and shaking continued for another minute. 960 mg of the powdered cell lysates was thawed and clarified by two successive centrifugations at 4 °C, 13,500*g* for 5 and 10 min, respectively, yielding ~400 μl of clarified lysate of which 380 μl (equivalent to ~650 mg of the original wet cell mass) was loaded onto two sucrose gradients in 190 μl aliquots.

Polysome isolation and digestion. Complexes sedimenting at 80 Svedberg units (S) or faster (including monosomes and polysomes, but excluding individual ribosomal subunits) were collected in the pellet fraction after centrifugation through a 10% to 20% sucrose gradient. 10% and 20% (w/v) sucrose stocks were prepared in buffer B (50 mM Tris-HCl pH 7.0 at 25 °C, 50 mM NH_4Cl, 4 mM $MgCl_2$, 1 mM DTT), and 0.7 ml of the 20% sucrose in buffer B was added to a 34 × 11 mm (~2.2 ml) thin walled polyallomer tube (Beckman-Coulter). The tube was then topped up with 10% sucrose in buffer B, sealed with Parafilm, tilted to 80° from the vertical and rotated about its axis for 90 s at 20 r.p.m., then slowly returned to vertical. The topmost 0.4 ml was removed and the tube rested for 1 h at 4 °C. Sample (up to 200 μl) was layered on top and tubes were centrifuged in a TLS-55 rotor (Beckman-Coulter) at 4 °C, 52,000 r.p.m. (179,740*g* average, k-factor 55.7) for 85 min. This centrifugation time and speed were optimized beforehand by visualizing the SSU, LSU and ribosome bands using GelStar nucleic acid stain (Lonza) in a series of experiments with different centrifugation conditions. The pellet was then rinsed in cold buffer A and resuspended by intense pipetting and scraping in 200 μl of buffer A supplemented with 1 mM DTT and 1× Complete EDTA-free Mini Protease Inhibitor (Roche). The resultant material was transferred into a 1.5 ml microcentrifuge tube, brought to the bottom by brief centrifugation and thoroughly homogenized using a microfuge pestle followed by vortexing at 4 °C for 2 min. Insoluble material was pelleted by centrifugation at 4 °C, 13,500*g* for 5 min and the clarified supernatant solution collected. A 15 μl aliquot of the solution was reserved for the Total Input library construction. The residual solution was digested with 15 U of *E. coli* RNase I (Ambion) per 1 OD_{260} unit of the solution at 23 °C for 45 min.

Separation of RNase-liberated polysomal SSU and ribosome complexes by sedimentation velocity. 7.5% to 30% w/v sucrose gradients were prepared with buffer B by adding 4 ml of 7.5% sucrose to a thin-walled polyallomer 14 × 89 mm ultracentrifuge tube (Beckman-Coulter) and underlaying it with 3.96 ml of 23% and then 3.3 ml of 30% w/v sucrose in buffer B. After topping up with 7.5% w/v sucrose in buffer B, the tube contents were mixed by rotation as for the 10–20% gradients (see above) and 1.5 ml was removed from the top. The RNase I-digested material (see polysome isolation and digestion section) was layered atop (occupying 5–10% of the tube length) and centrifuged in an SW 41 Ti rotor (Beckman-Coulter) at 4 °C, 38,000 r.p.m. (178,305*g* average, k-factor 143.9) for 4 h. The resultant gradients were fractionated at 1 ml min^{-1} and the SSU (~1.5 ml) and ribosome (~2 ml) fractions were identified by in-line monitoring of the absorbance profile at 254 nm and collected.

De-blocking and isolation of the crosslinked RNA. RNA was extracted and crosslinks reversed by the hot acid phenol procedure as previously described[33]. Typically, a 15 μl aliquot from the non-digested total input sample, a 50 μl aliquot of the ribosomes, or a 1,500 μl aliquot of the SSU gradient fractions was supplemented with 1% w/v sodium dodecyl sulphate, 10 mM EDTA, 10 mM Tris-HCl (pH 7.4 at 25 °C), 10 mM glycine and deionised water (if necessary) to result in 400 μl aliquots. 400 μl of 25:24:1 phenol:chloroform:isoamyl alcohol (pH ~4.5, Sigma-Aldrich) was then added, and the mixture shaken at 65 °C, 1,300 r.p.m. in a thermomixer for 45 min to de-crosslink. The mixture was then centrifuged at 15,000*g* at room temperature for 5 min, and the aqueous phase removed and nucleic acids precipitated from it by adding 20 μg of glycogen, 0.1 volumes of 3 M sodium acetate and 2.5 volumes of absolute ethanol. After incubation at −20 °C for at least 3 h, RNA was pelleted at 16,000*g* for 25 min, the RNA pellet washed twice in 80% ethanol and dried at 37 °C for at least 30 min in a sterile hood to evaporate residual phenol. Dried RNA was thoroughly resuspended by repeated pipetting and vortexing in 1 mM sodium citrate, pH 6.0, and assessed by measuring the absorbance spectrum.

Construction of the RNA sequencing libraries. One ribosomal and one SSU RNA sequencing library was constructed for the wild-type and the *SYO1*-TAP mutant strains, using a low-bias strategy generally as outlined before[34,35]. RNA concentration was estimated using a Bioanalyzer RNA Pico Chip (Agilent). De-crosslinked RNA (8 pmol) was heated for 2 min at 70 °C, immediately transferred to ice for 5 min, and then end-repaired with 20 U of 3′-phosphatase positive bacteriophage T4 polynucleotide kinase (PNK) (New England Biolabs), using conditions recommended by the supplier (1× PNK buffer without ATP, 1.5 U μl^{-1} RNaseOUT RNase inhibitor (Life Technologies), incubation for 2 h at 37 °C). The reaction was stopped and PNK inactivated by addition of EDTA to 1 mM and heating of the reaction mixture for 10 min at 65 °C, respectively. The end-repaired RNA was next 3′ polyadenylated with 2 U of *E. coli* poly(A) polymerase (EPAP) (Ambion) generally according to the supplier recommendations (1× EPAP buffer, 2.5 mM $MnCl_2$, 1 mM ATP, 1 mM DTT, and 0.8 U μl^{-1} RNaseOUT RNase inhibitor (Life Technologies), incubation for 1 h at 37 °C). The resulting polyadenylated RNA was ethanol precipitated as before, dried and dissolved in 10 mM HEPES-KOH (pH 7.6 at 25 °C), 0.1 mM EDTA. The RNA was next reverse transcribed using the split adaptor oligo(dT) primer (5′-phosphate-GA TCG TCG GAC TGT AGA ACT CTG AAC G /9-carbon spacer/ G TGA CTG GAG TTC CTT GGC ACC CGA GAA TTC CAT TTT TTT TTT TTT TTT TTT TVN-3′) with SuperScript III reverse transcriptase (Life Technologies), generally as recommended by the supplier. The RNA (up to 4 pmol) was first mixed with 20 pmol of the split adaptor primer, 10 nmol each of dNTPs and annealed by heating to 75 °C for 3 min, then cooling to 65 °C, and slow ramping to 55 °C. The reaction mixture was then supplemented with 5 mM DTT, 1 × First Strand Buffer (Life Technologies) and 2 U μl^{-1} RNaseOUT RNase inhibitor (Life Technologies) while heated, and then slow-ramped to 50 °C and incubated for 30 min. To purify the resultant cDNA away from the excess of the split adaptor primer, the reaction mixture was heated to 60 °C for 5 min, snap-cooled to 37 °C, supplemented with 20 U of *E. coli* exonuclease I (New England Biolabs) and incubated for 20 min. The reaction was then stopped by the addition of 23 mM EDTA. The residual non-extended split adaptor primer was removed by purification of the cDNA using Agencourt AMPure XP purification system (Beckman Coulter), strictly according to the manufacturer's recommendations, except for performing binding steps and removal of the supernatant with unbound material at 60 °C. Washing of the beads and elution of the cDNA with 10 mM HEPES-KOH (pH 7.6 at 25 °C), 0.1 mM EDTA were performed at room temperature. The resultant purified cDNA was further liberated from RNA contaminants by heating at 80 °C for 10 min, snap-chilling on ice, supplementing with 1 μl of RNase A/T mix (ThermoFisher Scientific), and incubating for 20 min at 37 °C. The purified cDNA was circularized with 50 U of CircLigase II (Epicentre/Illumina) generally according to the manufacturer's instructions (1× CircLigase II Buffer (Epicentre/Illumina), 2.5 mM $MnCl_2$, 500 mM betaine, incubation at 60 °C for 5 h). The reaction was stopped by the addition of 25 mM EDTA and heating at 80 °C for 10 min, and the circularised cDNA was purified using Agencourt AMPure XP purification system (Beckman Coulter) according to the manufacturer's recommendations. The cDNA was eluted with 10 mM HEPES-KOH (pH 7.6 at 25 °C), 0.1 mM EDTA and treated with Probe-Directed Degradation (PDD), as described[34,35]. The purified circularised cDNA was supplemented with 2 pmol of each of the depletion DNA probes (Supplementary Table 3)[35] and 1× Duplex-Specific Nuclease (DSN) buffer (Evrogen), and heated for 3 min at 95 °C, then cooled to 75 °C, and then slow ramped to the hybridization temperature of 48 °C. The mixture, while at the hybridization temperature, was then supplemented with 0.5 U of Duplex-Specific Nuclease (Evrogen) and further incubated for 20 min. The reaction was stopped by addition of 25 mM EDTA, the mixture diluted with sufficient amount of deionized water, phenol-chloroform (pH 8.0) extracted and gel-filtered using Micro Bio-Spin P-6 (Bio-Rad) columns. The PDD only partly removed rRNA content due to the unavoidable rRNA fragmentation and small depletion probe set employed. The resultant purified PDD-treated cDNA was amplified with Platinum *Pfx* DNA polymerase (Life Technologies), generally according to the manufacturer's

instructions, from primer pairs compatible with TrueSeq Small RNA Sample Preparation Kit (Illumina), bearing a unique tag for each library. The forward primers were 5′-CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX GTG ACT GGA GTT CCT TGG CAC CCG AGA ATT CCA-3′ (in which XXX XXX represents Illumina's indexing hexanucleotide sequences), and the reverse primer was 5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GA-3′. The amplification reaction included the purified PDD-treated cDNA, 0.3 mM each of dNTPs, 1× *Pfx* buffer (Life Technologies), 1.25 mM $MgSO_4$, 0.3 μM each of primers, 4.2 ng μl^{-1} extreme thermostable single-stranded DNA binding protein (New England Biolabs) and 0.05 U μl^{-1} Platinum *Pfx* DNA Polymerase (Life Technologies), and was typically repeated for 12–14 cycles using 94 °C melting, 62 °C (first two cycles) or 76 °C (third and higher cycles) annealing, and 68 °C polymerization temperatures. The amplified DNA samples from different libraries were equalized, pooled together, electrophoretically separated in a native agarose gel, selected for the size range of the insertion fragment of 10–250 nt, eluted from the gel by freezing the gel pieces at −20 °C for 2 h and recovering the DNA solution by immediate centrifugation of the gel pieces through 0.45 μm cellulose acetate filter at 13,000*g*. The recovered DNA solution was then phenol-chloroform (pH 8.0) extracted, ethanol precipitated (as described above), dissolved in 10 mM HEPES-KOH (pH 7.6 at 25 °C), 0.1 mM EDTA, quality-controlled with capillary electrophoresis (using Agilent High Sensitivity DNA chips), and directed to the high-throughput sequencing input[34].

The total input library was prepared as previously described[35] from the de-crosslinked RNA by a similar procedure, except that RiboMinus Eukaryote Kit (Life Technologies) was used to deplete the ribosomal RNA at the RNA stage and the RNA was fragmented with NEBNext RNA fragmentation module (New England Biolabs), instead of *E. coli* RNase I. Specifically, after crosslink reversal of the total input fractions (as described above), ethanol precipitation and re-solubilization of the total input RNA, the RNA was treated with RiboMinus Eukaryote rRNA removal kit (Life Technologies) strictly according to the manufacturer's instructions. The resultant rRNA-depleted total input RNA was fragmented with NEBNext Mg^{2+} RNA fragmentation module (New England Biolabs) at 94°C for 6 min, and next used for library construction in parallel to the ribosome and SSU samples as described above.

Sequencing was performed on a HiSeq 2500 (Illumina) using settings compatible with TrueSeq Small RNA Sample Preparation Kit (Illumina) and single-end reads of 150 nt.

Filtering and mapping of the RNA sequencing reads. The 150 nt Illumina reads were subjected to a quality trim using Trimmomatic (window: 7 nt; phred quality cutoff: 24) followed by adaptor trimming including the $(A)_{20}$ tract from the reverse transcription primer. The reads containing insert sequences without 3′ poly(A) and those shorter than 17 nt were discarded. To determine RNA biotype makeup of the libraries (Fig. 1c), the resultant insert sequences were then aligned to the *S. cerevisiae* genome (SacCer3) using Bowtie 2 (ref. 36). For all other analyses, a stepwise alignment strategy was employed as follows. Reads were first filtered to keep only those not mapping to the rRNA locus (chromosome XII, base pairs from 450,000 to 491,000). Remaining reads were then mapped to a custom tRNA sequence set and alignments were used to analyse tRNA content (see details below). Unmapped reads from this alignment were then aligned with spliced RNA sequences containing 'misc_RNA', 'ncRNA', 'tRNA' or 'snoRNA' primary tags in SacCer3. Finally, the remaining unmapped sequences were aligned to an mRNA reference consisting of all spliced sequences annotated as protein-coding mRNAs, plus 1,000 bp of flanking genomic sequence, allowing multiple mappings. Alignment coverage 'spikes' consisting of identical reads and that exceeded the read density for the surrounding transcript region by more than 20-fold were flagged as mapping artefacts and excluded. In cases when the 3′ end position of the trimmed read sequences was aligned immediately upstream of adenine nucleotides, it was impossible to resolve the exact location of the 3′ end of the corresponding original RNA fragment before polyadenylation. With these reads, the 3′ end position was randomly assigned with equal probability between all possible locations. The metagene analysis was robust to subsequent re-randomization: the proportions of read 3′ ends that changed by more than 2 nt were 2.26%, 1.51% and 2.20% for input, ribosome and SSU samples respectively, causing imperceptible changes at the metagene level.

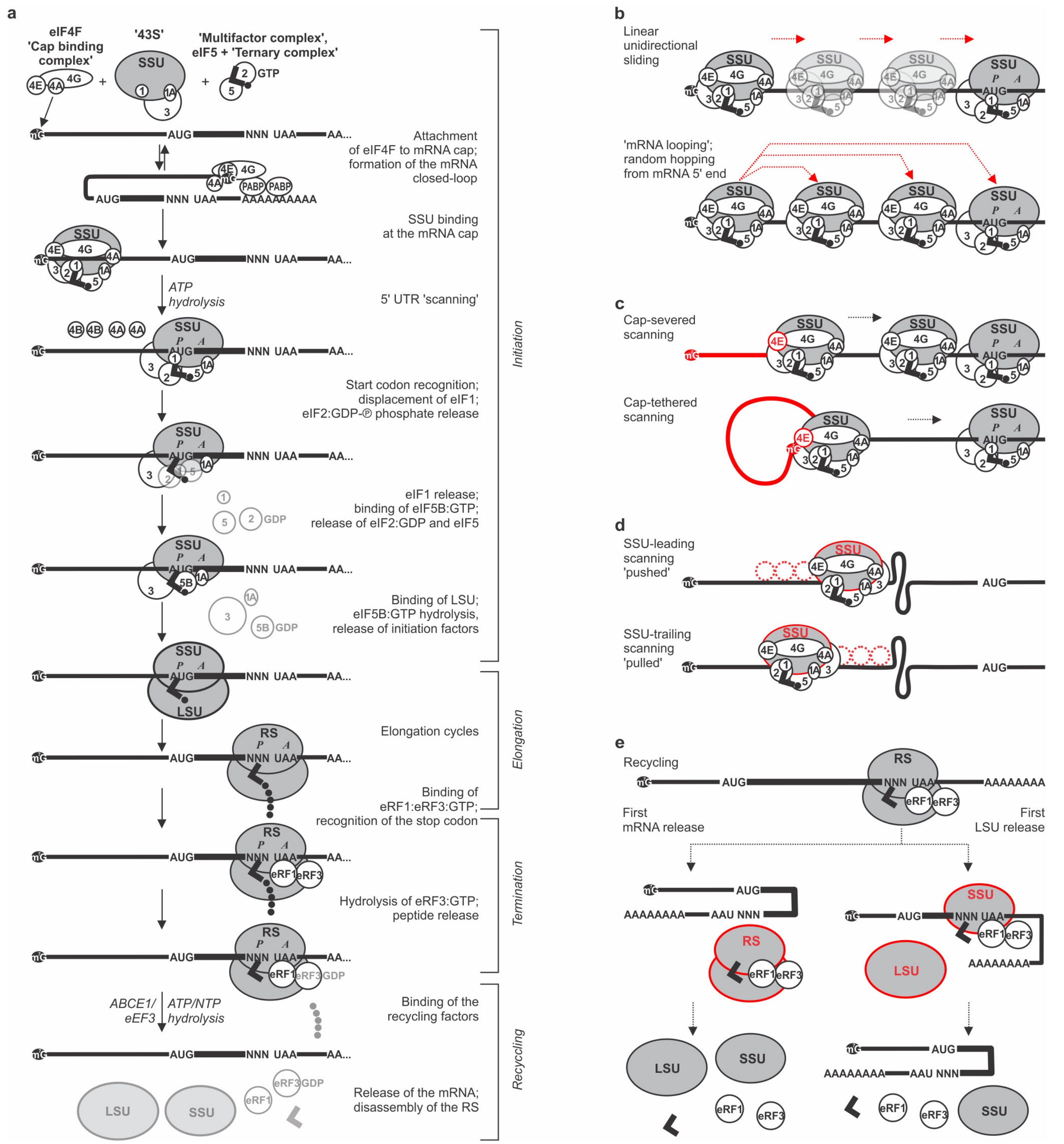
mRNAs with 5′ UTR introns without precise splicing coordinates[37] were removed from further analysis. We analysed skewness caused by library preparation by calculating the location-wise Shannon entropy (using the 'entropy' package in R) after aggregating input fraction fragment start/end position in the 14 most highly abundant ORFs (which had sufficiently few zero-read positions), treating each nucleotide position as a separate bin regardless of the sequence. All 14 ORFs exhibited Shannon entropies of >94% of that of a simulated (random, equiprobable) set of read positions, indicating low positional bias from the library preparation procedure. The RUST package[38] was also used to analyse skewness in the ribosome footprints, using an A-site offset of 17 and footprints of 31 nt in length, and to compare with such in known low-bias yeast ribosome profiling data sets[39,40]. Due to an observed effect of library size on RUST scores, the two published data sets analysed were first randomly subsampled to equalize the library size. A trough in footprints with 3′ ends located over uridyl nucleotide residues of AUG start codons (detected as a diagonal line in, for example, Fig. 2a, Extended Data Fig. 5a, or as a vertical line in, for example, Extended Data Fig. 6a, c, to the left from the shorter of the SSU peaks) was associated with a small bias against 3′-terminal U's and AU dinucleotides, as confirmed by analysing the aggregate densities of ORF-associated footprint 3′ ends located over internal AUG triplets and out-of-frame stop codons (Extended Data Fig. 4b–d).

Metagene footprint end-densities were also subjected to gene-level bootstrapping to estimate the influence of individual genes on the density profile. Genes were randomly re-selected, allowing duplications, *n* times in which *n* was the original number of genes (5,859). Footprint end counts for each nucleotide within the regions of interest (−110 to +150 nt relative to start codon, and −170 to +80 nt relative the last ORF sense codon) were generated and normalized to total reads in the region of interest. This process was repeated 1,000 times and the mean and 95th percentiles of the footprint-end density at each nucleotide position were calculated.

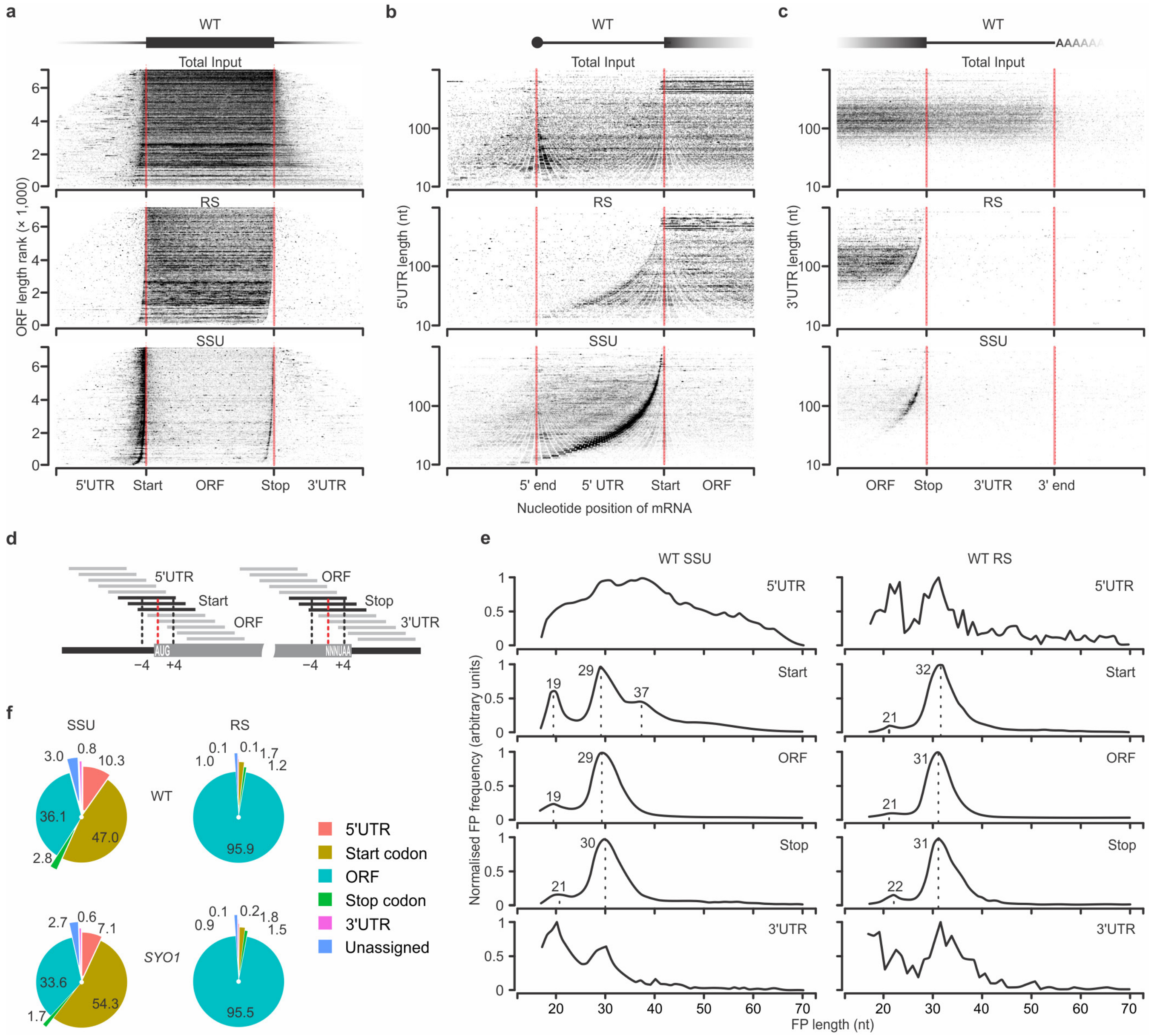
For tRNA content analysis, to overcome the multitude of similar isoforms and non-templated addition of 'CCA' to the 3′ ends of tRNAs, a custom reference was constructed. Non-redundant tRNA sequences were downloaded from GtRNAdb[41] (S288c strain, April 2011) and 'CCA' sequences added to the 3′ ends if these were not template-specified. Only insert sequences mapping definitively to one tRNA were counted and results from the wild-type and *SYO1*-TAP lines were treated as duplicates for the tRNA analysis in edgeR[42].

Code Availability. Custom R scripts were used for this and all other analyses, available from the authors on request.

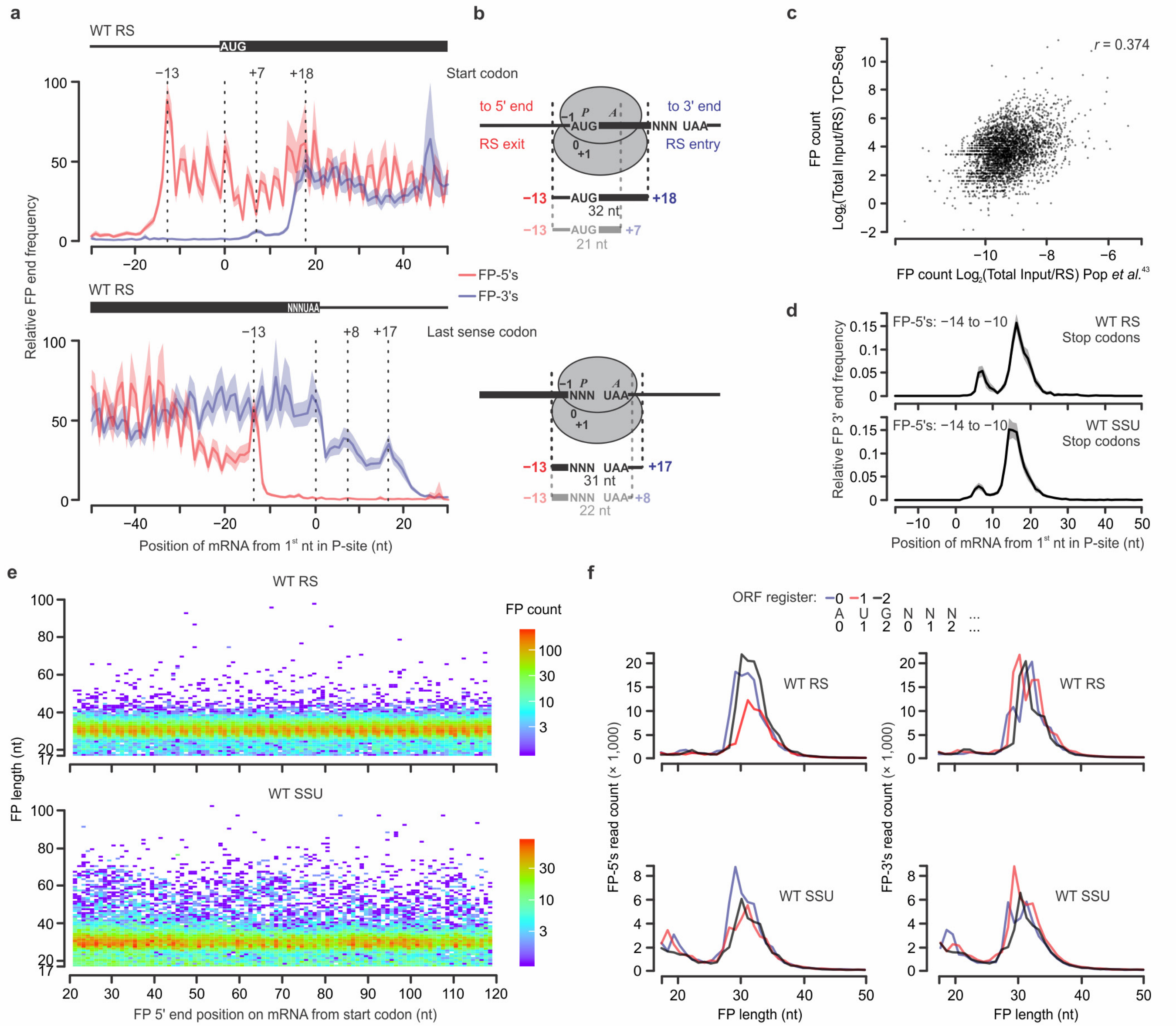
31. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425,** 737–741 (2003).
32. Kressler, D. *et al.* Synchronizing nuclear import of ribosomal proteins with ribosome assembly. *Science* **338,** 666–671 (2012).
33. Archer, S. K., Shirokikh, N. E., Hallwirth, C. V., Beilharz, T. H. & Preiss, T. Probing the closed-loop model of mRNA translation in living cells. *RNA Biol.* **12,** 248–254 (2015).
34. Archer, S. K., Shirokikh, N. E. & Preiss, T. Probe-directed degradation (PDD) for flexible removal of unwanted cDNA sequences from RNA-seq libraries. *Curr. Protoc. Hum. Genet.* **85,** 11 15 11–11 15 36 (2015).
35. Archer, S. K., Shirokikh, N. E. & Preiss, T. Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. *BMC Genomics* **15,** 401 (2014).
36. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
37. Zhang, Z., Hesselberth, J. R. & Fields, S. Genome-wide identification of spliced introns using a tiling microarray. *Genome Res.* **17,** 503–509 (2007).
38. O'Connor, P., Andreev, D. & Baranov, P. Surveying the relative impact of mRNA features on local ribosome profiling read density in 28 datasets. *bioRxiv* http://dx.doi.org/10.1101/018762 (2015).
39. Artieri, C. G. & Fraser, H. B. Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.* **24,** 2011–2021 (2014).
40. McManus, C. J., May, G. E., Spealman, P. & Shteyman, A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res.* **24,** 422–430 (2014).
41. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37,** D93–D97 (2009).
42. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
43. Pop, C. *et al.* Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.* **10,** 770 (2014).
44. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
45. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320,** 1344–1349 (2008).



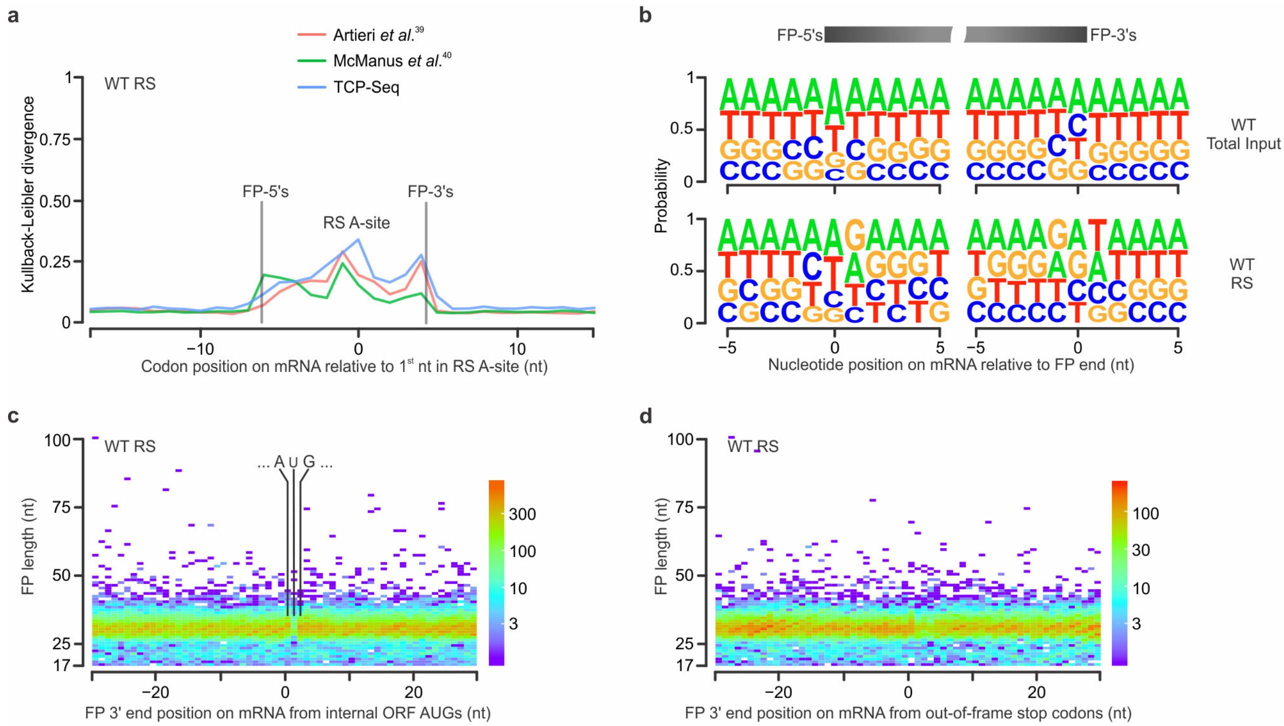
Extended Data Figure 1 | Current model of eukaryotic translation with emphasis on substeps with scarce, missing or ambiguous evidence. **a**, Schematic representing the current model of the translation cycle starting from SSU attachment to mRNA (top) and ending with recycling of the post-termination complexes on mRNA (bottom). Each step represents complexes with existing *in vivo* or *in vitro* evidence. **b**–**e**, Alternative models suggested for different substeps, with key differences highlighted in red. **b**, Type of scanning motion to locate the cognate start codon: linear unidirectional motion of SSU along the 5′ UTR (top) or random hopping over 5′ UTR regions from mRNA 5′-end (bottom). **c**, Mode of SSU attachment during scanning: cap-severed mRNA scanning in which initiating SSU lose their interaction with the 5′ cap structure as they encroach into the inner 5′ UTR regions (top), as opposed to cap-tethered scanning where the SSU 5′-end cap connection is preserved at least until start codon recognition. **d**, Position of the scanning factors relative to the SSU: SSU-led scanning in which initiation factors responsible for the scanning motion are located and interact with mRNA behind the SSU to push it in 5′→3′ direction (top), or an SSU-trailing mechanism where these factors interact with mRNA in front of the moving SSU to pull it along. **e**, Alternative recycling of termination complexes: mRNA is released from terminating ribosomes before they are disassembled into subunits (left), in contrast to a staged recycling where the LSU is first ejected from the post-termination ribosomes, leaving post-termination SSU attached to mRNA (right).



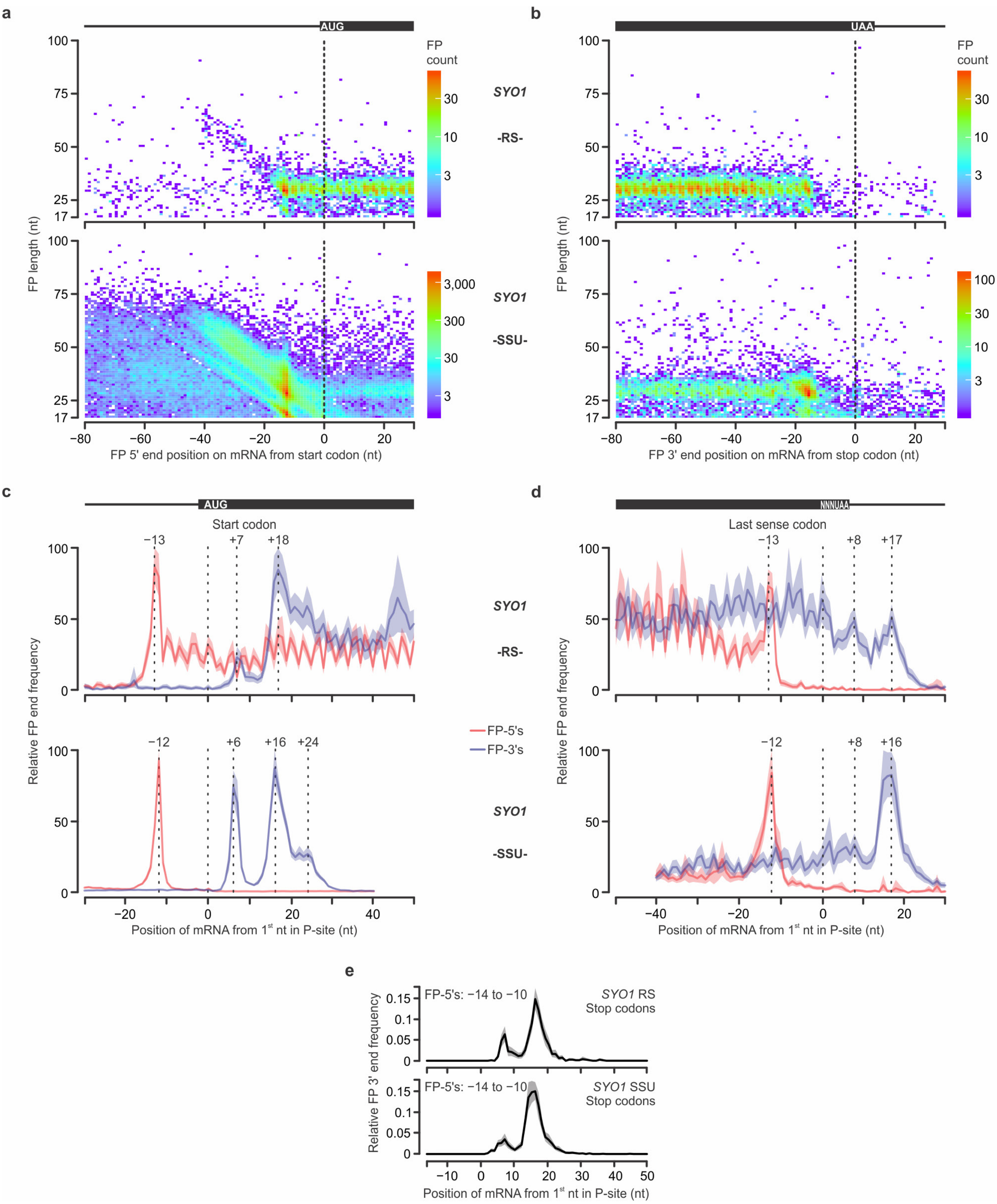
Extended Data Figure 2 | Overview of the *in vivo* distribution of translation complexes on mRNAs measured by TCP-seq. a, Snapshots of translation as represented by footprint 5′ ends on mRNA from wild-type strain ribosomes and SSU fractions (middle and bottom panels), total input mRNA is shown for reference (top panel). Footprint reads are mapped to all annotated mRNAs up to 1,000 nt of mRNA 5′ and 3′ flanking regions. Footprint 5′ end positions are plotted (horizontal axis) against the ORF length rank (vertical axis). Horizontal location (both within and outside the ORF) is scaled relative to ORF length, with start and stop codon positions indicated at the first nucleotide of these triplets (red lines). Footprints with 5′ ends accumulating at a constant distance upstream from start and stop appear as curved vertical trails due to the scaling. Transparency of points is increased by $\log_{10}(2 +$ read number per transcript) to allow the distribution in less abundant transcripts to be observed. **b**, **c**, Same as **a**, but the horizontal location is scaled to 5′ (**b**) or 3′ UTR (**c**) length, and footprint 5′ ends are plotted against 5′ or 3′UTR length, respectively. **d**, Scheme showing classification of the footprints based on position along mRNAs for **e**, **f**. **e**, Profiles of the frequency of the footprint lengths in the 5′ UTR, start codon, ORF, stop codon and 3′ UTR regions for the ribosome and SSU fractions of the wild-type yeast. Only footprints mapping to mRNAs within annotated UTRs are analysed. The maximum frequency value for each plot is scaled to 1. **f**, Proportions of the footprints in the different mRNA regions for the ribosome and SSU libraries of wild-type and *SYO1* yeast strains. For raw read counts, see Supplementary Table 1.



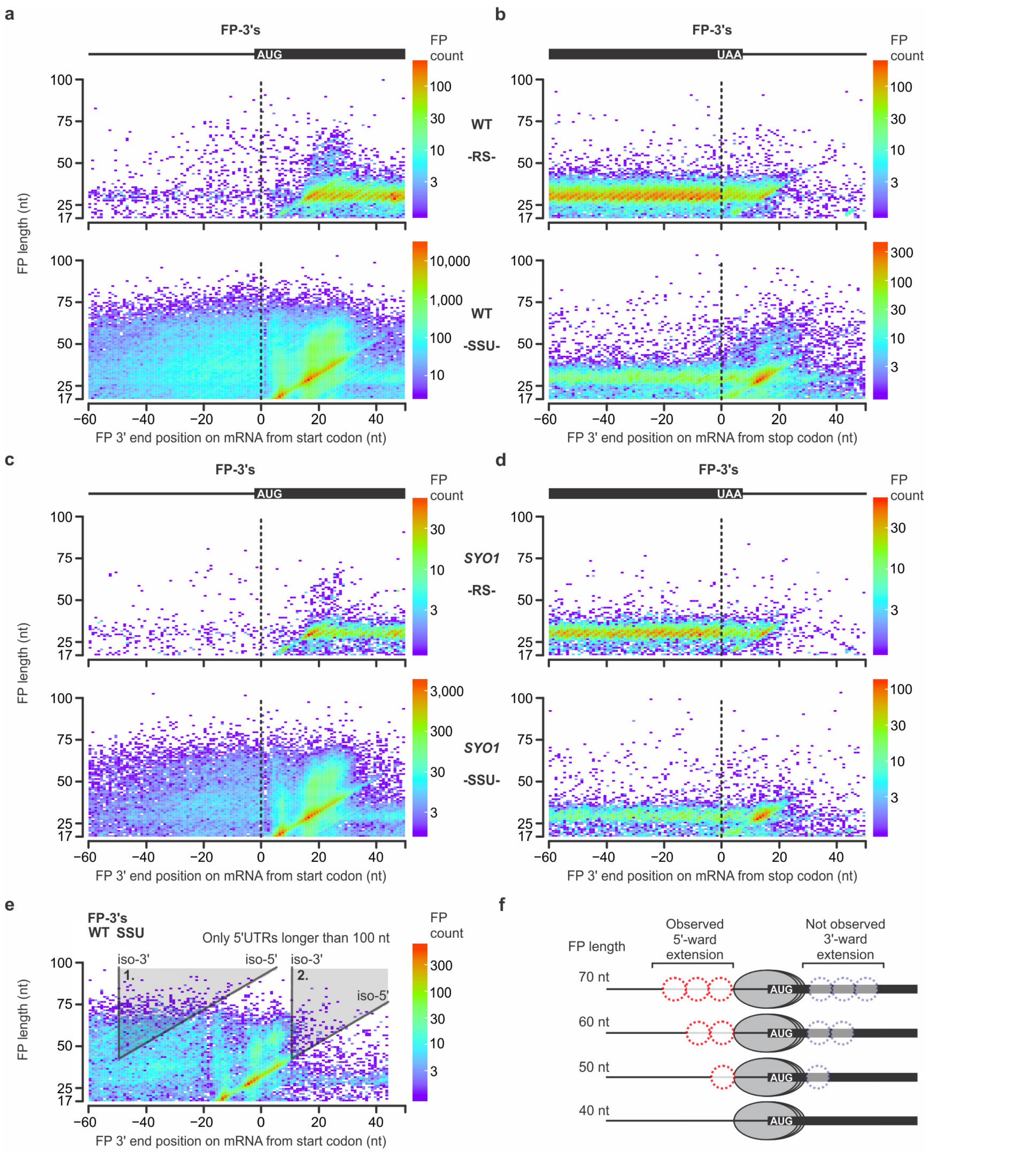
Extended Data Figure 3 | Characteristics of start-codon-associated, stop-codon-associated and ORF-associated ribosome footprints and comparison of TCP-seq with ribosome profiling. a, Aggregated densities of wild-type yeast footprint 5′ ends (FP-5′s; red) and footprint 3′ ends (FP-3′s; blue) from the ribosome fraction, mapped to all annotated mRNAs and shown within the vicinity of the start (top) or end (bottom) of the ORFs (aligned as in Fig. 4a, c). Positions of the major footprint end density modalities are indicated (dashed lines). The error regions shown as shading of the same colour indicate the 95th percentiles of 1,000 iterations of random gene-wise resampling (see Methods for details). **b**, Schematics illustrating the calculated position and 3′ and 5′ footprint extensions for the major ribosome footprint types (alignment as for the SSU in Fig. 4b, d) across the start (top) or stop (bottom) codons. **c**, Correlation of the ribosome density on mRNAs in exponentially propagating wild-type yeast as measured by TCP-seq (vertical axis) or translation (ribosome) profiling[43] (horizontal axis). Note that total input denotes either total polysomal RNA (TCP-seq) or total cellular mRNA (ribosome profiling). **d**, End density for ribosome footprint 3′ ends plotted relative to the last sense codon from all ORFs, only for footprints with 5′ ends within the main peak for terminating ribosome complexes (from −14 nt to −10 nt). Error regions are depicted as in **a**. **e**, Same as in Fig. 2a (top panel), but showing ORF region from +20 to +120 nt from the start codon. **f**, Register bias (triplet periodicity) for 5′ and 3′ ends of ORF-associated SSU and ribosome footprints. Footprints falling between +10 and +200 nt from the start codons of all mRNAs are binned into one of three registers according to the remainder when dividing the footprint 5′ or 3′ ends position (relative to the ORF start) by three. The relative difference between the frequency of the most common and least common register is 16.5% for ribosome and 9.9% for SSU. The SSU register is shifted 1 nt downstream to that of the ribosome, possibly due to a 1 nt recession in the footprint 5′ ends after LSU dissociation.



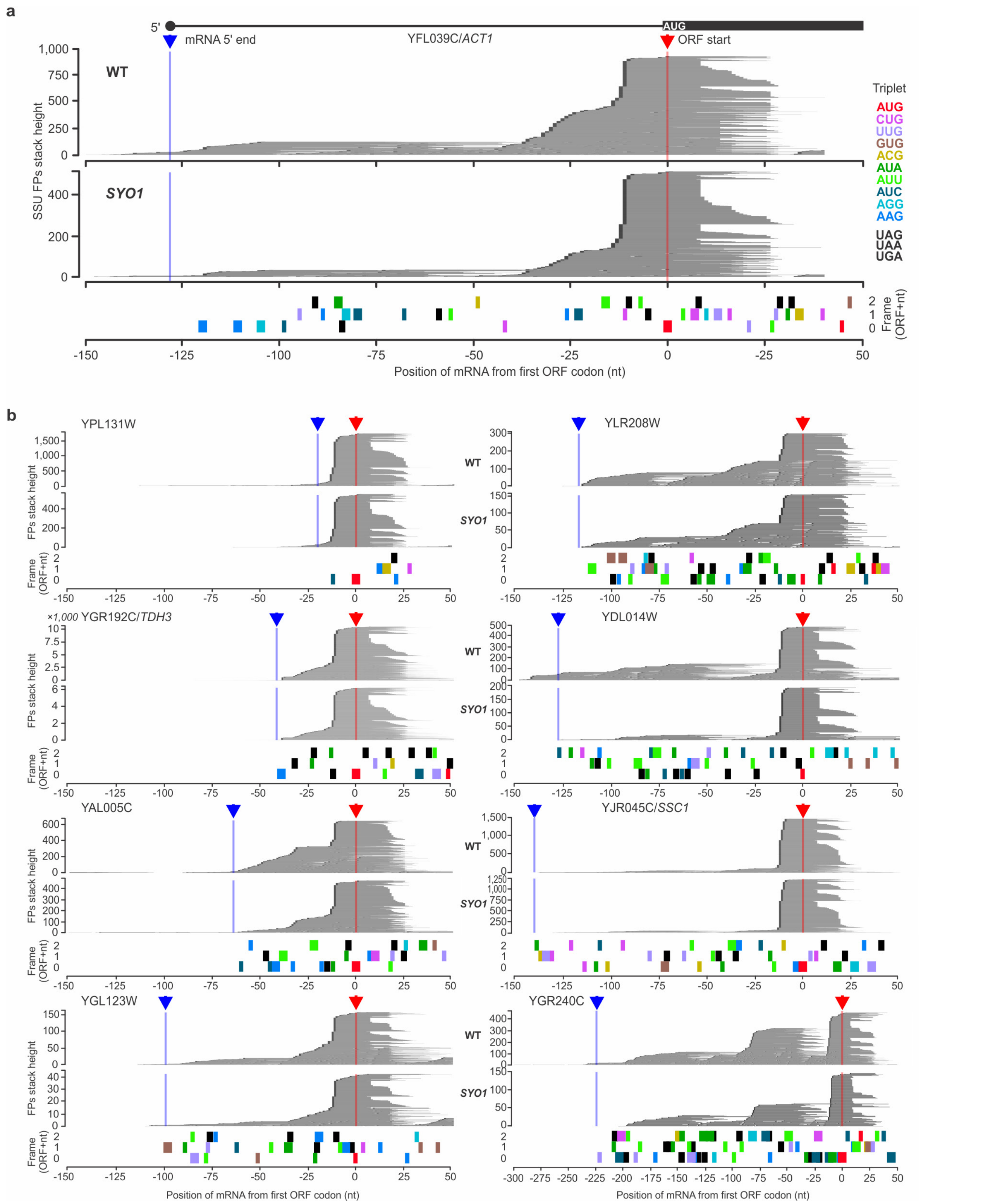
Extended Data Figure 4 | Assessment of potential library or sequence bias. a, Comparison of sequence bias in the TCP-seq ribosome fraction to other ribosomal profiling experiments in yeast[39,40], using codon-level Kullbach–Liebler divergence as implemented in the RUST package[38]. Peaks in skewness at the fragment ends likely result from library preparation artefacts while the peak at the central A-site (+17 nt from the footprint 5′ end) probably results from biological differences in codon decoding speed[38]. **b**, Probabilities for occurrence of nucleotides around the 5′ or 3′ ends of TCP-seq reads (footprint 5′ and 3′ ends) generated using Weblogo 3 (ref. 44), **c**, **d**, A depletion of footprint 3′ ends at the uridyl nucleotides. **c**, Depletion of footprint 3′ ends over U of the AUG triplets is characteristic of all AUG triplets within ORFs, and not only for start AUG codons (visible for example, in Extended Data Fig. 6a, c). Internal AUG triplets were selected from 50 nt downstream of start codons up to ORF ends, and footprint 3′ end densities aligned by these codons. **d**, Same as **c**, but footprint 3′ ends were aligned around the first nucleotides of out-of-frame stop codons located within ORF regions.



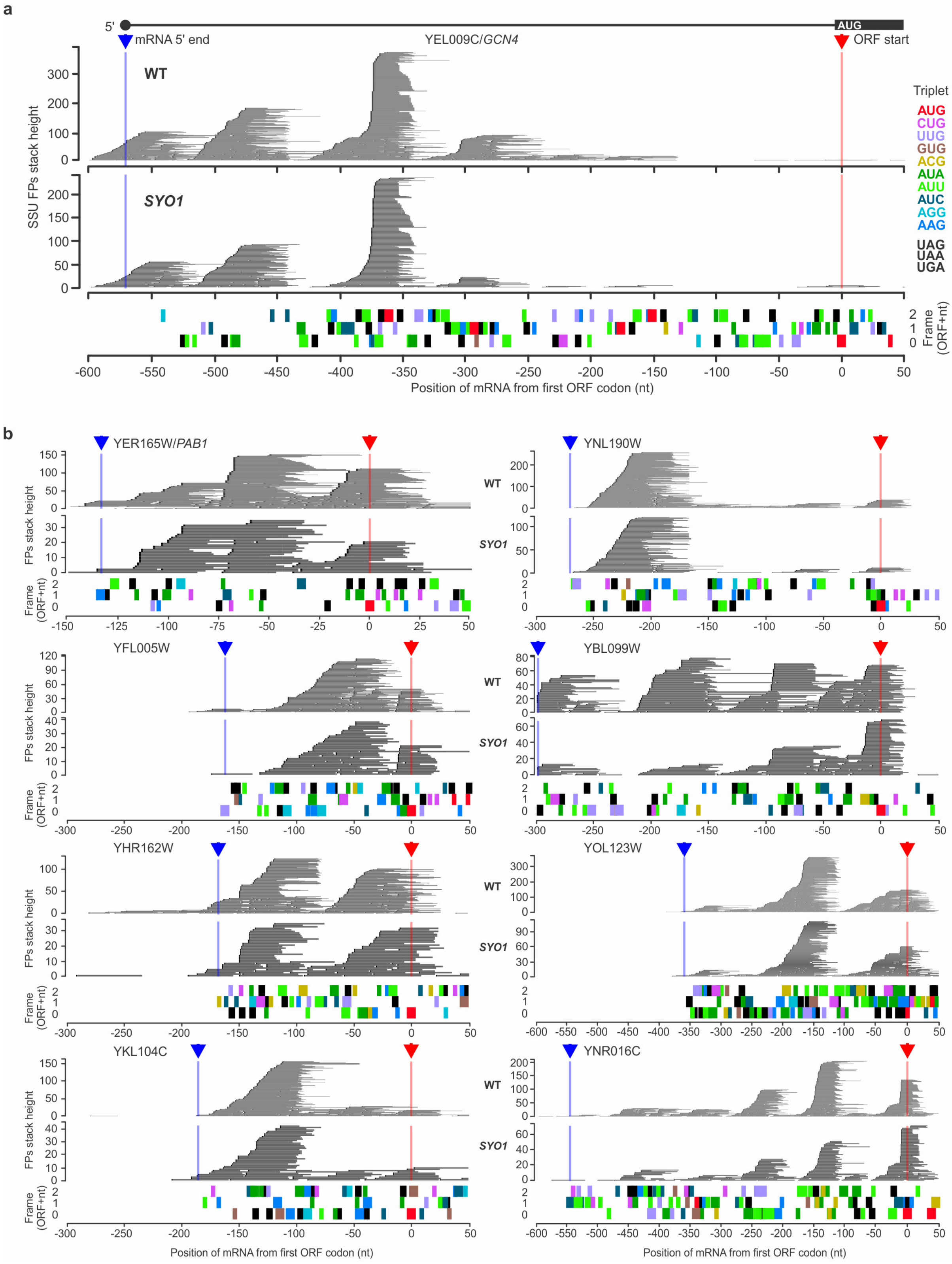
Extended Data Figure 5 | Assorted analyses of TCP-seq data for the *SYO1* yeast strain with mild deficiency in cytoplasmic LSU levels[11]. **a**, **b**, As in Fig. 2a, b. **c**, **d**, Top panels as in Extended Data Fig. 3a, bottom panels as in Fig. 4a, c. **e**, As in Extended Data Fig. 3d, but for the *SYO1* cells. Note that horizontal axes in **c**, **d** are aligned by the position of the first nucleotide in the P-site of initiating (left panels) or terminating (right panels) to allow for consistency in footprint 5′ and 3′ end extensions across translation complexes. For terminating complexes, this corresponds to the first nucleotide of the last sense codon (not the stop codon). Error regions in **c**–**e** indicate the 95th percentiles of 1,000 iterations of gene-wise bootstrapping (see Methods).



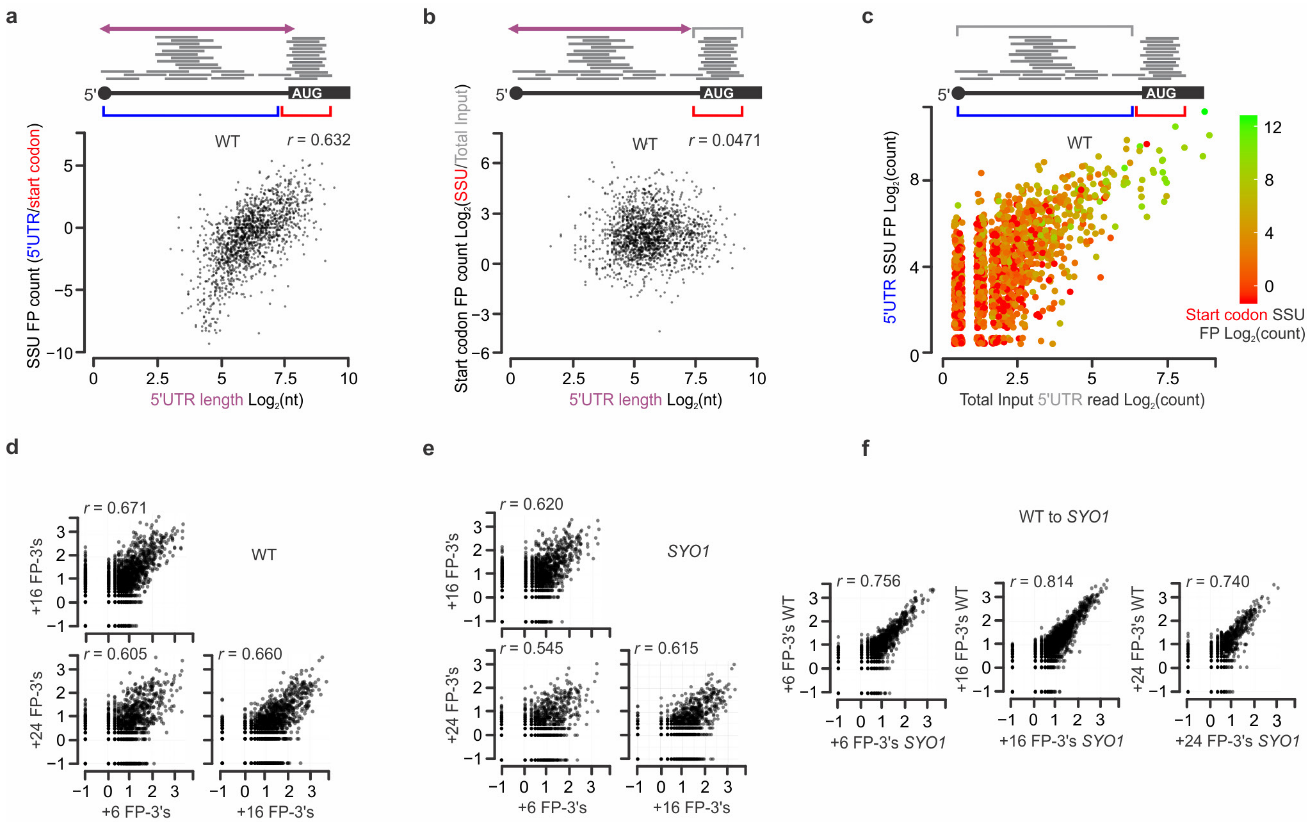
Extended Data Figure 6 | Snapshots of SSU and ribosome footprint distribution reveal translation initiation by a cap-severed, SSU-led scanning mechanism and staged ribosome recycling after termination, visualized as footprint length versus footprint 3′ end positions aggregated across all mRNAs. a, **b**, As in Fig. 2a, b. **c**, **d**, As in Extended Data Fig. 5a, b, but for footprint 3′ ends. **e**, Heatmap of footprint 3′ end positions relative to start codons, aggregated amongst mRNAs having an annotated 5′ UTR of at least 100 nt in length[45]. The extended scanning SSU are arrested when the leading edge (footprint 3′ end) overruns the start codon by a set margin, irrespective of the trailing edge (footprint 5′ end) position. This is evidenced by the scarcity of longer footprints with 3′ end > +30 nt position and 5′ end < −13 nt position (window 2; encompassing 175 footprints) relative to that expected from an equivalently bounded region (footprint 3′ end > −30 nt and footprint 5′ end < −73 nt) that is fully within the 5′ UTR (window 1; encompassing 1,552 footprints). Isoline iso-3′ and iso-5′ margins forming window 1 and 2 represent positions of the footprint 3′ end and footprint 5′ end isolines of SSU located upstream of (window 1) or on (window 2) the start codons. **f**, Schematic of plausible topologies of scanning SSU complexes as they approach and align over the start codons, to assist in interpreting data in **e**.



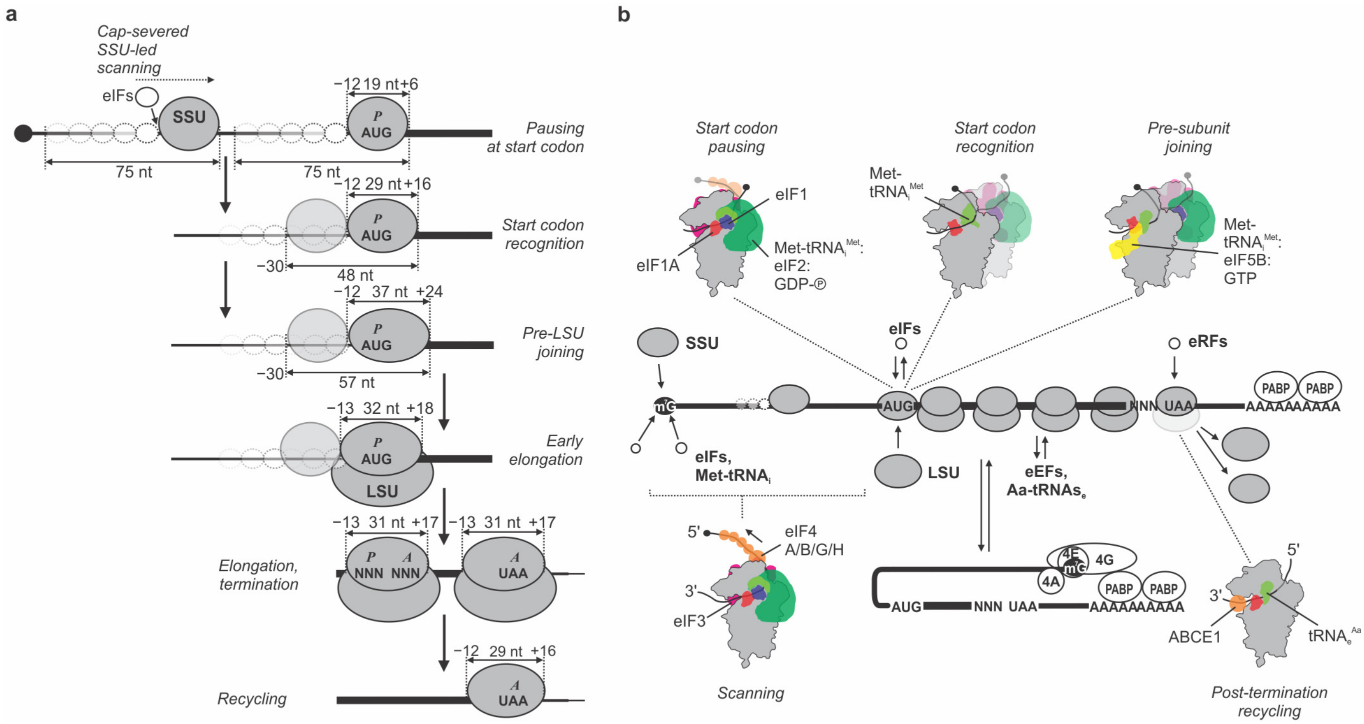
Extended Data Figure 7 | TCP-seq data for individual mRNAs from wild-type and *SYO1* cells demonstrating uninterrupted 5′ UTR scanning and high frequency of initiation. a, Translation initiation on the 5′UTR of *ACT1* mRNA coding for actin. Individual SSU footprints observed on the *ACT1* mRNA of wild-type (upper panel) and *SYO1* (middle panel) cells. Footprint 5′ ends of each footprint are shown in black. Annotated mRNA 5′ end[45] and ORF start are indicated by blue and red arrows above the panels and lines of the same colour; potential start codons and stop codons along the 5′ UTR are indicated below in each of three reading frames (lower panel), colour-coded by bars (with the legend on the right). Initiation codons located in perfect nucleotide context (for yeast) are shown as bars of double width. **b**, Same as **a**, but for a collection of additional mRNAs. See also the web interface (http://dx.doi.org/10.6084/m9.figshare.3206725) for browsing and analysing the TCP-seq data for wild-type and *SYO1* yeast strains, available at (http://bioapps.erc.monash.edu/TCP/).



Extended Data Figure 8 | TCP-seq data for individual mRNAs from wild-type and *SYO1*-TAP cells demonstrating SSU pile-ups within 5′ UTRs and decreased frequency of initiation. Designations and abbreviations as in Extended Data Fig. 7. **a**, Translation initiation on the 5′ UTR of *GCN4* mRNA coding for general control nonderepressible bZIP transcriptional activator of amino acid biosynthetic genes. **b**, Same as **a**, but for a collection of other mRNAs. See also the web interface for browsing and analysing the TCP-seq data for wild-type and *SYO1* yeast strains (http://bioapps.erc.monash.edu/TCP/).



Extended Data Figure 9 | Efficiency of scanning and start codon selection across yeast mRNAs. a, Dependency of the ratio of 5′ UTR to start codon-associated SSU footprints on 5′ UTR length. **b**, Dependency of the relative SSU count at start codons on the length of the 5′ UTR. **c**, Dependency of 5′ UTR SSU footprint counts (vertical axis) on the abundance and 5′ UTR length of mRNA (horizontal axis; as determined by 5′ UTR RNA fragment counts in the total input fraction) and translation initiation rates (colour; as determined by count of start-codon occupying SSU footprints). Data shown in **a**–**c** is for the wild-type yeast strain. **d**, **e**, Cross-correlation of the three major start codon footprint types by 3′ end position (refer to Fig. 4a, b) observed across mRNAs of the wild-type (WT) (**d**) and *SYO1*-TAP (*SYO1*) (**e**) strains. **f**, Correlation between wild-type and *SYO1* yeast cells of the gene-wise footprint count of each of the three major start codon footprint types. Pearson's correlation coefficients are indicated (r). All counts were incremented by 0.5 to allow logarithmic transformation.



Extended Data Figure 10 | Summary of TCP-seq findings and their interpretation in the context of existing mechanistic models and information on ribosome structure. **a**, Schematic depicting the proposed progression of translation initiation, including cap-severed, SSU-led scanning, and structural rearrangements at start codons. Post-termination events featuring lingering of SSU at stop codons are also shown. **b**, View of the translation cycle with emphasis on the nexus between TCP-seq footprint types and recent structural information on the SSU in complex with translation factors. The schematic models represent structures obtained for *in vitro* assembled complexes that are taken to represent particular sub-steps of the translation cycle. Note that i*n vivo* translation intermediates might not completely match the composition and/or structural configuration of *in vitro* assembled complexes. Schematic models representing the anticipated arrangement of scanning and start codon pausing SSU[14,19–22], start codon recognition[14,19–22], pre-subunit joining[23,24] and post-termination recycling[25] SSU complexes are shown. eIF1 (blue), eIF1A (red), eIF2 (dark green), initiator $tRNA_i^{Met}$ (light green), eIF3 (purple), eIF4A/B/G/H and ABCE1 (amber), eIF5B (yellow), and elongator $tRNA_e^{Aa}$ (light green).

LETTER

doi:10.1038/nature18589

Structural organization of the inactive X chromosome in the mouse

Luca Giorgetti[1]*†, Bryan R. Lajoie[2]*, Ava C. Carter[3]*, Mikael Attia[1]*, Ye Zhan[2], Jin Xu[3], Chong Jian Chen[1], Noam Kaplan[2], Howard Y. Chang[3], Edith Heard[1,4] & Job Dekker[2,5]

X-chromosome inactivation (XCI) involves major reorganization of the X chromosome as it becomes silent and heterochromatic. During female mammalian development, XCI is triggered by upregulation of the non-coding *Xist* RNA from one of the two X chromosomes. *Xist* coats the chromosome in *cis* and induces silencing of almost all genes via its A-repeat region[1,2], although some genes (constitutive escapees) avoid silencing in most cell types, and others (facultative escapees) escape XCI only in specific contexts[3]. A role for *Xist* in organizing the inactive X (Xi) chromosome has been proposed[4–6]. Recent chromosome conformation capture approaches have revealed global loss of local structure on the Xi chromosome and formation of large mega-domains, separated by a region containing the *DXZ4* macrosatellite[7–10]. However, the molecular architecture of the Xi chromosome, in both the silent and expressed regions, remains unclear. Here we investigate the structure, chromatin accessibility and expression status of the mouse Xi chromosome in highly polymorphic clonal neural progenitors (NPCs) and embryonic stem cells. We demonstrate a crucial role for *Xist* and the *DXZ4*-containing boundary in shaping Xi chromosome structure using allele-specific genome-wide chromosome conformation capture (Hi-C) analysis, an assay for transposase-accessible chromatin with high throughput sequencing (ATAC–seq) and RNA sequencing. Deletion of the boundary disrupts mega-domain formation, and induction of *Xist* RNA initiates formation of the boundary and the loss of DNA accessibility. We also show that in NPCs, the Xi chromosome lacks active/inactive compartments and topologically associating domains (TADs), except around genes that escape XCI. Escapee gene clusters display TAD-like structures and retain DNA accessibility at promoter-proximal and CTCF-binding sites. Furthermore, altered patterns of facultative escape genes in different neural progenitor clones are associated with the presence of different TAD-like structures after XCI. These findings suggest a key role for transcription and CTCF in the formation of TADs in the context of the Xi chromosome in neural progenitors.

To investigate the structure of the Xi chromosome, we performed allele-specific Hi-C in a clonal neural progenitor cell (NPC) line that was derived from highly polymorphic F_1 mouse embryonic stem (ES) cells (129 × Cast, Extended Data Fig. 1a–f). We first performed Hi-C in ES cells, in which XCI has not yet occurred, and found that autosomes and both active X chromosomes displayed prominent active/inactive (A/B) compartmentalization and TAD structures (Extended Data Figs 2a–c and 3). In NPCs, compartments and TADs were similarly detected on autosomes and the active X (Xa) chromosome (Fig. 1a, Extended Data Fig. 3a). Notably, however, the Xi chromosome displayed no A/B compartments (Extended Data Fig. 2c), but was instead partitioned into two massive interaction domains separated by a hinge region of ∼200 kb including the *DXZ4* macrosatellite[11,12] (Fig. 1a), as recently reported for the human and mouse Xi chromosomes[7–9]. Furthermore, TADs were found to be largely absent on the Xi chromosome (Fig. 1a), as previously suggested[9,10].

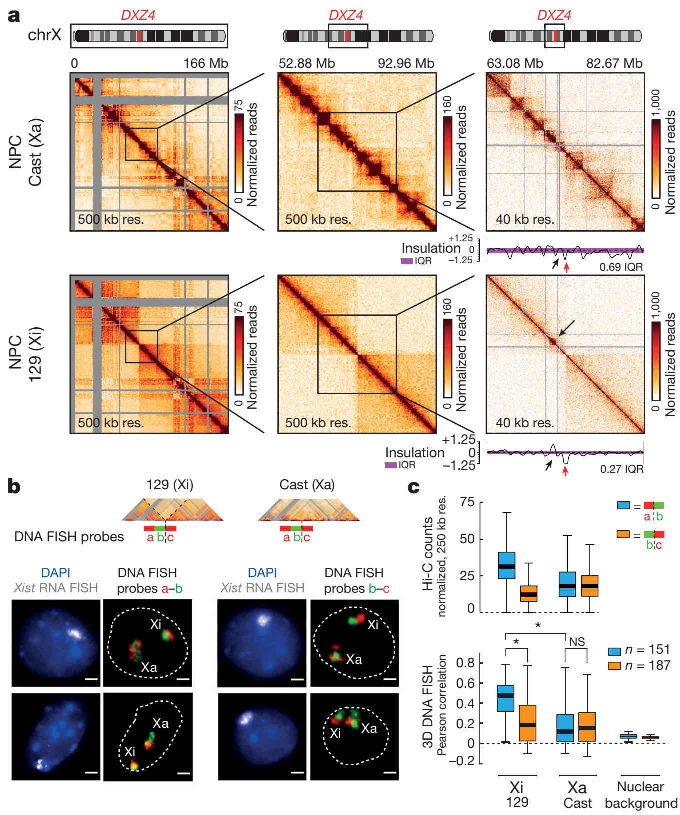


Figure 1 | The distinct conformation of the Xi and Xa chromosomes. **a**, Allele-specific Hi-C maps of the Xa and Xi chromosomes in NPCs (left), and two increasingly smaller regions centred around *DXZ4* (centre and right). Purple areas in the insulation score plots indicate the interquantile range (IQR) of insulation scores over the entire X chromosome, to illustrate reduced insulation scores along the Xi chromosome (indicating loss of TADs). Black arrow: position of a residual TAD. Red arrow: position of *DXZ4*. **b**, Top, scheme of DNA FISH probes. Bottom, probes a–b (within the same mega-domain) are more overlapping and spherical on the Xi than on the Xa chromosome, whereas signals from b–c (across the mega-domain boundary) show partitioning on the Xi chromosome into two separate domains. **c**, Loci detected by probes a–b are more interacting than b–c on the Xi chromosome, in both Hi-C (top) and 3D-DNA FISH (bottom). *$P < 8 \times 10^{-17}$ (Wilcoxon's rank sum test corrected with Bonferroni). NS, not significant. *n* denotes number of cells analysed in DNA FISH. Centre lines: medians, all experiments were performed in biological duplicates.

[1]Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, 26 Rue d'Ulm, 75248 Paris Cedex 05, France. [2]Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, Massachusetts 01605, USA. [3]Center for Personal Dynamic Regulomes and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. [4]Collège de France, 11 place Marcelin-Berthelot, Paris 75005, France. [5]Howard Hughes Medical Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester, Massachusetts 01605, USA. †Present address: Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland.
*These authors contributed equally to this work.

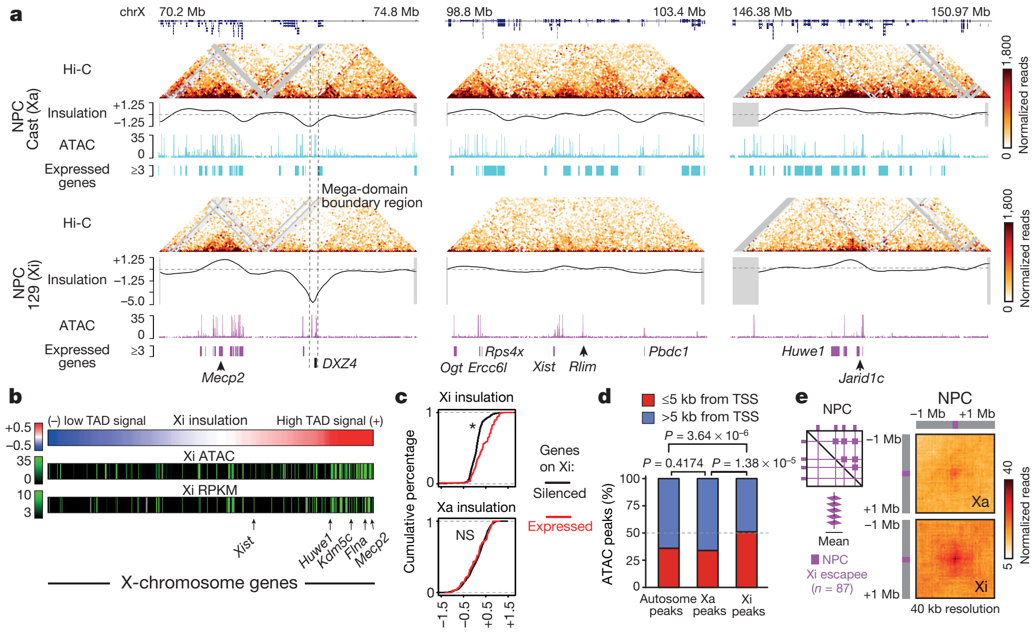


Figure 2 | Expression, chromatin accessibility and chromatin conformation along the Xi chromosome. a, Hi-C interactions, insulation score, ATAC–seq signal and location of expressed genes (via RNA-seq) in three regions of the Xa and Xi chromosome. Left to right: the cluster of facultative escape genes containing *Mecp2*, the *Xist* locus and the region encompassing *Jarid1c*. Hi-C data are shown at 40-kb resolution. **b**, Regions with increased TAD structure harbour promoters that are expressed and accessible on the Xi chromosome, as shown by analysis of Hi-C insulation (TAD structure), ATAC–seq read counts, and RNA-seq reads per kilobase of transcript per million mapped reads (RPKM). Each column represents a promoter, sorted by insulation score in the 40-kb interval containing the promoter, lowest-to-highest (weakest-to-strongest TAD signal). ATAC counts are extracted from promoters (TSS ± 500 base pairs (bp)). **c**, Xi-chromosome-expressed genes (escapees) fall within regions with higher insulation scores on the Xi chromosome as compared to Xi-chromosome-silenced genes ($*P = 4.44 \times 10^{-16}$, Fisher exact test test), despite having similar insulation scores on the Xa chromosome ($P = 0.43114$, Kolmogorov–Smirnov test). **d**, ATAC–seq peaks on the Xi chromosome tend to be closer to TSSs (within 5 kb) than peaks on autosomes and the Xa chromosome (statistical significance was assessed by Fisher exact test). **e**, Interaction pile-up maps showing mean Hi-C signal for all pairwise combinations of the 87 wild-type NPC Xi chromosome escapees on the both the Xa and the Xi chromosome.

To investigate the spatial organization of mega-domains at the single-cell level, we performed DNA fluorescence *in situ* hybridization (FISH) with 18-Mb probe sets located within one mega-domain (probes a–b), or spanning the mega-domain boundary (probes b–c) (Fig. 1b, Extended Data Fig. 4a). Despite extensive cell-to-cell variation, regions within the same mega-domain showed greater overlap on the Xi than on the Xa chromosome (Fig. 1b, c, Extended Data Fig. 4b), whereas regions spanning the boundary showed lower overlap on the Xi chromosome, in agreement with Hi-C data (Fig. 1c). Similar results were obtained in an NPC clone with a Cast Xi chromosome and in astrocytes (Extended Data Fig. 4c). Using two independent quantification methods (Extended Data Fig. 4d–f), the volume of each 18-Mb region was found to be modestly (approximately 20%) but significantly smaller on the Xi than the Xa chromosome, consistent with observations on the human Xi chromosome[13–15]. Thus, the mouse Xi chromosome is moderately compacted and partitioned into two large, spatially distinct domains that show varying degrees of overlap within the cell population (Fig. 1b, c, Extended Data Fig. 4).

Although the Xi chromosome globally presented no TADs in NPCs, we could detect a few residual TAD-like structures (Fig. 1a, black arrow). Integration of Hi-C, RNA sequencing (RNA-seq)[16] and ATAC–seq data produced in the same lines revealed that these Xi chromosome TAD-like structures correspond to hotspots of residual transcription and open chromatin. The RNA-seq and ATAC–seq profiles are consistent with global inactivity of the Xi chromosome (Extended Data Fig. 5a), with a 75% reduction in both the number of expressed genes and accessible elements compared to the Xa chromosome. Most ATAC–seq peaks on the Xi chromosome fall in the pseudoautosomal region, the *Xist* locus and at the promoters of genes that escape XCI either facultatively or constitutively[3] (Extended Data Fig. 5a). The amount of local structure on the Xi chromosome correlates with the density of transcribed loci and accessible elements (Extended Data Fig. 5d), as shown in Fig. 2a by three examples: a cluster of 19 facultative escapees including *Mecp2*, overlapping a strong ~800-kb TAD-like structure; the *Xist* locus with moderate interactions across a ~250-kb region; and 5 escapees including the constitutive *Jarid1c* (also known as *Kdm5c*) and the facultative *Huwe1* genes, embedded in a ~500-kb TAD. Importantly, residual TAD-like structures on the Xi chromosome occasionally coincide with sub-TAD structures on the Xa chromosome (Fig. 2a). Generally, escapees are located in Xi chromosome regions with higher TAD strength and chromatin accessibility as compared to silenced genes (Fig. 2b, c, Extended Data Fig. 5b). Thus, in NPCs, the appearance of TAD structures is intimately linked to gene expression on the Xi chromosome, unlike on the Xa chromosome and autosomes, where TADs are present even in the absence of transcription.

Notably, 51% of Xi chromosome accessible sites are <5 kb from a promoter compared to ~35% on the Xa chromosome (Fig. 2d), suggesting that escape is often regulated through promoter-proximal sites. Most ATAC–seq peaks on the Xi chromosome were found at CTCF-binding sites (Extended Data Fig. 5c), implicating CTCF in escape. Recent findings suggest that cohesin (which co-localizes with CTCF) is globally lost on the Xi chromosome, which may lead to the chromosome-wide loss of TADs[9]. Our discovery that only escapees show TAD-like structures on the Xi chromosome, and that they are associated with putative CTCF sites, is consistent with a role for CTCF in TAD formation and/or maintenance[17–19].

Escapees on the Xi chromosome also tend to interact with each other even across the mega-domain boundary, consistent with previous circularized chromosome conformation capture (4C) analysis results[5] (Fig. 2e and Extended Data Fig. 2d, e).

To investigate the importance of the unusual bipartite organization of the Xi chromosome we deleted the ~200 kb boundary region encompassing the *DXZ4* locus, specifically on the 129 allele in ES cells (ΔFT) (Extended Data Figs 1b and 6a). After differentiation, many NPC clones were isolated with a 129 (ΔFT) Xi chromosome. Deletion of the boundary did not affect XCI establishment, as NPC clones with either a wild-type or ΔFT Xi chromosome were obtained. Hi-C performed on one such clone (D9B2) revealed massive reorganization of the ΔFT Xi chromosome resulting in fusion of the two mega-domains (Fig. 3a). No effect was visible on the Cast Xa chromosome (Extended Data Fig. 7a). DNA FISH confirmed that sequences on either side of the deleted boundary overlap significantly more on the ΔFT Xi chromosome compared to wild type, consistent with Hi-C data (Extended Data Fig. 8a).

Notably, in the D9B2 clone, facultative escapees (*Mecp2*, *Huwe1*) no longer escaped from XCI on the ΔFT Xi chromosome (Extended Data Fig. 6b-c). ATAC–seq and RNA-seq confirmed substantially reduced escape on the ΔFT Xi chromosome, with only 29 expressed genes (Extended Data Fig. 8b). Transcription and chromatin accessibility were lost at 66 of the 87 facultative escape genes (~76%) (Fig. 3b), but maintained at all 6 constitutive escapees (Fig. 3b, Supplementary Table 1 and Extended Data Fig. 8c). However, analysis of multiple NPC clones, both wild type and ΔFT, revealed only a slight tendency for ΔFT NPC clones to show less escape than wild type, with varying degrees of facultative escape between clones, with or without the mega-domain boundary (Extended Data Fig. 6d). Nevertheless, clone D9B2 provided a unique opportunity to study the relationship between transcriptional activity and chromosome conformation on the Xi chromosome.

TAD-like structures were absent on the Xi chromosome when expression was lost at facultative escapees in the D9B2 clone (Fig. 3b and Extended Data Fig. 7b). Xi chromosome-wide comparisons between the D9B2 (ΔFT) and wild-type NPCs revealed a strong correlation between loss of escape, loss of chromatin accessibility and reduction in TAD signal (Fig. 3c). Specific long-range interactions between escapees were also lost on the ΔFT Xi chromosome (Fig. 3d). Eight genes showed *de novo* escape in the D9B2 clone (Supplementary Table 1). These *de novo* escapees were not clustered or highly accessible by ATAC–seq, and their expression did not lead to strong TAD formation (data not shown).

Of the 224 Xi chromosome ATAC–seq peaks that we detected in the wild-type clone, 139 were lost in the D9B2 clone (Fig. 3e, Extended Data Fig. 8b). These lost sites were enriched for promoter-proximal location (64%) (Fig. 3f, g). In total, 93% of these promoter-proximal sites contain CTCF-binding sites, an enrichment compared to the 64% of promoter-distal sites (Fig. 3f). These CTCF sites are closer to escape gene transcription start sites (TSSs) than unchanging sites, again pointing to a role for promoter-proximal CTCF in escape gene regulation (Fig. 3h).

We next investigated the role of *Xist* in establishing the unusual organization of the Xi chromosome. We induced *Xist* expression in undifferentiated male (XY) ES cells carrying a tetracycline-inducible promoter at the endogenous *Xist* locus[1]. Hi-C revealed that 48 h of *Xist* induction resulted in notable structural changes along the X chromosome, and these changes were not observed after induction of a Xist mutant lacking the A-repeat region (Fig. 4a, b)—which cannot silence genes (Extended Data Fig. 9a), but is competent for *Xist* coating and exclusion of RNA polymerase II (refs 1, 4). Wild-type *Xist* induction did not lead to detectable changes in TAD structure (data not shown), but resulted in increased interaction frequencies along the chromosome (Fig. 4b, d). The contact map of the wild-type *Xist*-coated X chromosome was found to be more similar to that of the NPC Xi chromosome than either the non-induced or the A-repeat mutant *Xist*-coated X chromosome (Fig. 4a). Notably, physical separation across the mega-domain boundary occurred (Fig. 4b), confirmed by RNA/DNA FISH (Fig. 4c, Extended Data Fig. 9b). Induction of wild-type

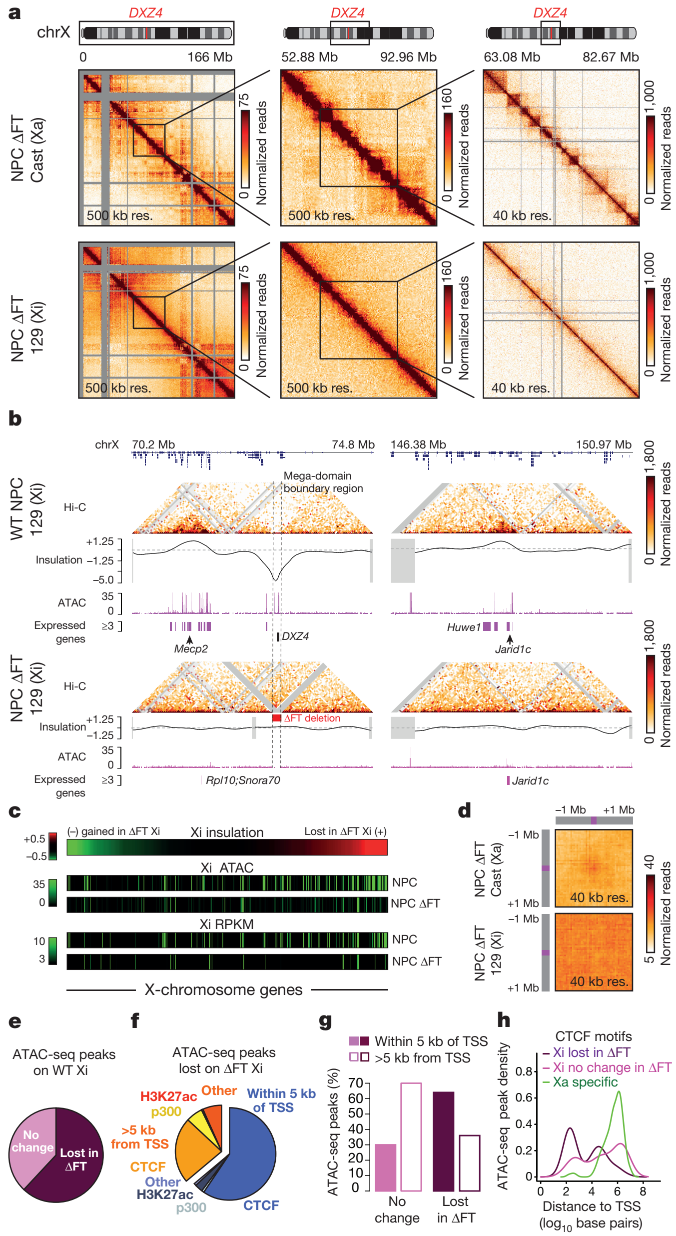


Figure 3 | Deletion of the mega-domain boundary leads to loss of bipartite folding. **a**, Hi-C contact maps for Xa (Cast) and Xi (129 ΔFT) chromosomes in mega-domain boundary mutant NPCs (left), and for increasingly smaller regions (centre and right). **b**, Two regions on the ΔFT Xi chromosome showing Hi-C, RNA-seq and ATAC–seq signal. The same regions as in Fig. 2a are shown. ATAC–seq and RNA-seq from the Xi chromosome in wild-type NPCs are included for reference (shown in Fig. 2a). **c**, Loss of TAD structure in ΔFT NPCs correlates with loss of accessibility (ATAC) and expression (RPKM). Each column represents a promoter. Heat maps are sorted by change in TAD strength (insulation score) from wild-type to ΔFT Xi in the 40-kb interval containing the promoter. ATAC counts are extracted from promoters (TSS ± 500 bp TSS). **d**, Interaction pile-up map showing mean interaction signal in ΔFT NPCs for all pairwise combinations of the 87 wild-type NPC Xi chromosome escapees. **e**, Quantification of ATAC–seq peaks in wild-type and ΔFT NPCs on the Xi chromosome. Of 224 Xi chromosome peaks in the wild type, 139 are lost in the mutant. **f**, Chromatin immunoprecipitation followed by sequencing (ChIP–seq) annotation of ATAC–seq peaks lost in ΔFT NPCs. **g**, Distance from TSS of ATAC–seq peaks that are lost, or do not change, after deletion of the mega-domain boundary. **h**, Distance from escapee TSS of CTCF peaks that are lost in ΔFT NPCs, do not change in ΔFT NPCs, and peaks on the Xa chromosome.

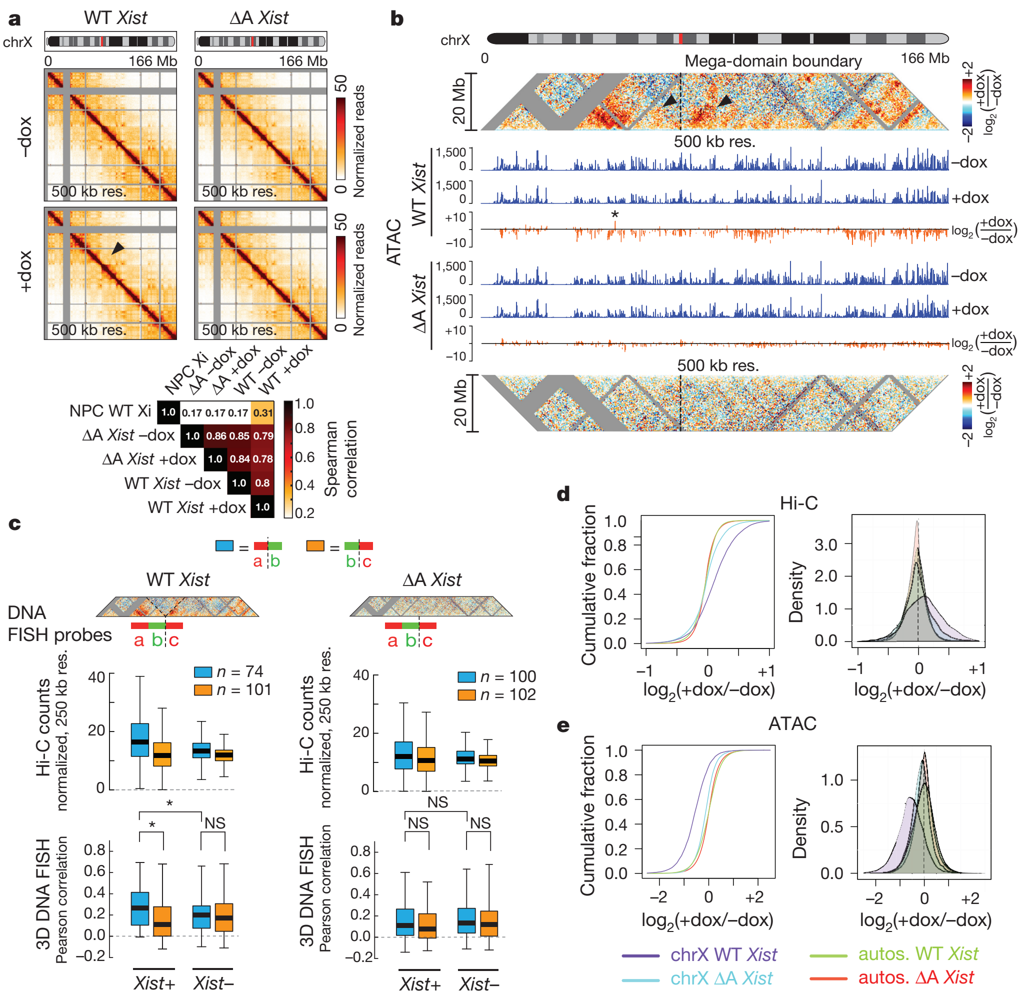


Figure 4 | *Xist*-mediated silencing is sufficient to generate a boundary at *DXZ4* in ES cells. a, Top, Hi-C analysis on the X chromosome in male ES cells expressing wild-type (WT) or A-repeat mutant (ΔA) *Xist*, before (−dox) and after (+dox) 48 h induction. dox, doxycycline. Bottom, correlation analysis of male X chromosome and NPC Xi chromosome Hi-C maps, showing increased similarity between the male *Xist*-coated X and Xi chromosomes after induction of wild-type, but not ΔA *Xist*. **b**, Alignment of structural changes detected in Hi-C and changes in chromatin accessibility measured by ATAC–seq after 48 h of wild-type or ΔA *Xist* expression. Dashed line: *DXZ4* position. Arrowheads: increased interactions on either side of the mega-domain boundary. Asterisk: genomic position of *Firre*. **c**, RNA/DNA FISH showing increased overlap of probes a–b on the *Xist*-coated X chromosome after 48 h of wild-type, but not ΔA *Xist* induction. Probes b–c show lower overlap (indistinguishable from non-*Xist* coated chromosomes in cells where *Xist* expression was not induced after doxycycline treatment), in line with Hi-C predictions (see Methods). *Xist*− signals correspond to the −dox sample. $*P < 0.05$ (Wilcoxon's rank sum test corrected with Bonferroni). *n* denotes number of cells analysed in DNA FISH. Centre lines: medians. Boxes: middle 50% of data points. Two biological replicates were analysed. **d**, Changes in Hi-C contact probability after 48 h of wild-type or ΔA *Xist* expression, indicating a wild-type *Xist*-specific increase in contact probability on the X chromosome. **e**, Changes in the ATAC–seq signal at X chromosome and autosomal peaks, indicating wild-type *Xist*-specific loss of chromatin accessibility on the X chromosome.

Xist from one X chromosome in female ES cells[20] generated a boundary of comparable magnitude (Extended Data Fig. 9c). ATAC–seq showed that 48 h of wild-type (but not A-repeat mutant) *Xist* expression results in globally reduced (but not eliminated) accessibility on the X chromosome (Fig. 4b, e). Interestingly, a small number of loci showed increased chromatin accessibility after wild-type *Xist* coating including the *Firre* long noncoding RNA (Fig. 4b), which together with the *DXZ4* macrosatellite has been proposed to anchor the Xi chromosome to the nucleolus[21]. Regions of reduced chromatin accessibility tend to show increased interaction frequencies with loci in the surrounding ~20 Mb of genomic sequence (Extended Data Fig. 9d). We note that conformational changes on the *Xist*-coated X chromosome in male ES cells were relatively mild compared to the NPC Xi chromosome, probably for several reasons. First, *Xist* was induced in ~35–45% of cells (in two independent replicates). Detecting structural alterations in such a sub-population may be difficult by Hi-C, and this may explain our observation that TADs are still detected in the overall cell population. Second, the *Xist*-coated Xi chromosome chromatin state in ES cells may not be comparable to that in differentiated NPCs[22].

In conclusion, our study uncovers a complex molecular architecture of the Xi chromosome, and reveals new insights into TAD formation. The Xi chromosome is moderately compacted and organized into two mega-domains, with global loss of TADs, except at clusters of expressed genes. Thus, in contrast with the notion that TADs are highly stable across differentiation and do not require transcription for their presence or maintenance[10,23], our findings demonstrate that (1) TADs can be lost in the context of the Xi chromosome, at least in NPCs; and (2) transcription together with binding of factors such as CTCF may enable their maintenance or *de novo* re-creation.

The bipartite folding of the Xi chromosome into two mega-domains is evolutionarily conserved, pointing to a possible role for this peculiar organization in the XCI process. Deletion of the boundary region, leading to fusion of the two mega-domains, does not affect initiation of XCI. However, somewhat reduced rates of facultative,

but not constitutive, escape are observed. Facultative escapees are silenced during XCI and then re-expressed[24,25] and may be more prone to stochastic events and/or the influence of long-range interactions with other parts of the X chromosome, such as the CTCF-rich *DXZ4* region[26]. As promoter-proximal CTCF sites characterize both facultative and constitutive escape genes, we speculate that transient interactions of escapees with the boundary region during XCI may influence escape (Extended Data Fig. 9e). However, the highly variable facultative escape that we found between NPC clones, both with or without the *DXZ4* boundary, suggests that escape efficiency cannot be due just to the presence of the *DXZ4* region, but may be influenced by local folding and accessibility to CTCF binding. The exact timing and mechanism of facultative escape, and its relationship with TAD formation, will require further investigation. In conclusion, our study establishes the Xi chromosome as a powerful system for studying the mechanistic relationships between chromosome conformation and gene regulation, and points to key roles for gene activity and CTCF in the establishment of TAD structure in the context of facultative heterochromatin.

Note added in proof: Consistent with our observations in mouse, deletion of *DXZ4* from the human Xi chromosome results in loss of mega-domain structure (E. Lieberman Aiden & B. Chadwick, Deletion of the macrosatellite *DXZ4* on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl Acad. Sci USA* (in press)).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 July 2015; accepted 18 May 2016.
Published online 18 July 2016.

1. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nature Genet.* **30,** 167–174 (2002).
2. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161,** 404–416 (2015).
3. Peeters, S. B., Cotton, A. M. & Brown, C. J. Variable escape from X-chromosome inactivation: identifying factors that tip the scales towards expression. *BioEssays* **36,** 746–756 (2014).
4. Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **20,** 2223–2237 (2006).
5. Splinter, E. *et al.* The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* **25,** 1371–1383 (2011).
6. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341,** 1237973 (2013).
7. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **162,** 687–688 (2015).
8. Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16,** 152 (2015).
9. Minajigi, A. *et al.* Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349,** aab2276 (2015).
10. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
11. Giacalone, J., Friedes, J. & Francke, U. A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nature Genet.* **1,** 137–143 (1992).
12. Chadwick, B. P. DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res.* **18,** 1259–1269 (2008).
13. Teller, K. *et al.* A top-down analysis of Xa- and Xi-territories reveals differences of higher order structure at ≥ 20 Mb genomic length scales. *Nucleus* **2,** 465–477 (2011).
14. Naughton, C., Sproul, D., Hamilton, C. & Gilbert, N. Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures. *Mol. Cell* **40,** 397–409 (2010).
15. Eils, R. *et al.* Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. *J. Cell Biol.* **135,** 1427–1440 (1996).
16. Gendrel, A.-V. *et al.* Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* **28,** 366–380 (2014).
17. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111,** 996–1001 (2014).
18. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32,** 3119–3129 (2013).
19. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports* **10,** 1297–1309 (2015).
20. Schulz, E. G. *et al.* The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* **14,** 203–216 (2014).
21. Yang, F. *et al.* The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* **16,** 52 (2015).
22. Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* **5,** 695–705 (2000).
23. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
24. Corbel, C., Diabangouaya, P., Gendrel, A.-V., Chow, J. C. & Heard, E. Unusual chromatin status and organization of the inactive X chromosome in murine trophoblast giant cells. *Development* **140,** 861–872 (2013).
25. Patrat, C. *et al.* Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proc. Natl Acad. Sci. USA* **106,** 5198–5203 (2009).
26. Horakova, A. H. *et al.* The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. *Genome Biol.* **13,** R70 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Heard, Dekker, and Chang laboratories for their help and critical insights; PICT-IBiSA@BDD (UMR3215/U934) Imaging facility of the Institut Curie. L.G. would like to thank L. Mirny for discussing gyration tensor analysis. Supported by grants from the National Institutes of Health (P50-HG007735) and Scleroderma Research Foundation (to H.Y.C.), from the National Human Genome Research Institute (R01 HG003143) and the National Institutes of Health Common Fund, National Institute of Diabetes and Digestive and Kidney Diseases (U54 DK107980) to J.D., the Human Frontier Science Program to N.K., an EMBO Fellowship to L.G., an ERC Advanced Investigator award (ERC-2014-AdG no. 671027), EU FP7 grants SYBOSS (EU 7th Framework G.A. no. 242129) and MODHEP (EU 7th Framework G.A. no. 259743), La Ligue, Fondation de France, Labex DEEP (ANR-11-LBX-0044) part of the IDEX Idex PSL (ANR-10-IDEX-0001-02 PSL) and ABS4NGS (ANR-11-BINF-0001) to E.H. J.D. is an investigator of the Howard Hughes Medical Institute.

Author Contributions E.H. and J.D. conceived the original strategy. For Hi-C, M.A. and L.G. prepared the ES cell and NPC samples, Y.Z. performed the Hi-C experiments; for ATAC–seq, H.Y.C., E.H. and J.D. designed the experiments; A.C.C. and J.X. prepared the samples and performed the experiments; for RNA-seq, M.A. and L.G. prepared the samples and performed the experiments and C.J.C. analysed the data. Integrated analysis of Hi-C sequencing, RNA-seq and ATAC–seq data was performed by B.R.L., J.X., L.G. and A.C.C., with assistance from C.J.C. and N.K. and input from J.D.; L.G. and E.H. designed the FISH experiments, L.G. and M.A. performed FISH experiment and L.G. analysed data. M.A., L.G. and E.H. designed the NPC and *DXZ4* mutant strategy, M.A. and L.G. performed the experiments and analysed them. L.G., B.R.L., A.C.C., E.H. and J.D. wrote the manuscript with input from H.Y.C.

Author Information Sequencing data have been deposited in the Gene Expression Omnibus (GEO) under accession numbers GSE72697 (Hi-C); GSE71156 (ATAC–seq); and GSE72697 (boundary deletion data). Code used in this work is available at (https://github.com/dekkerlab/giorgetti-nature-2016). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.H. (edith.heard@curie.fr) or J.D. (job.dekker@umassmed.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture. The hybrid mouse ES cell line F121.6 (129Sv-Cast/EiJ), a gift from J. Gribnau, was grown on mitomycin-C-inactivated mouse embryonic fibroblasts in ES cell media containing 15% FBS (Gibco), 10^{-4}M β-mercaptoethanol (Sigma), 1,000 U ml^{-1} leukaemia inhibitory factor (LIF, Chemicon). Male-inducible TXY and TXY:ΔA lines (a gift from A. Wutz, called *Xist-tetOP* and *Xist-ΔSX-tetOP*, respectively, in ref. 1) were cultured in the same conditions and treated for 48 h with 2 μg ml^{-1} doxycycline. Differentiation of F121.6 ES cells into NPCs was performed as previously described[16]. Subcloning of NPCs was made by limiting dilution and manual colony picking. All cells used in this study were characterized for absence of mycoplasma contamination.

Boundary deletion. To generate the boundary region deletion, 5×10^6 ES cells were transfected with 5 μg each of two plasmids (pX459) each expressing Cas9 and a chimaeric guide RNA (gRNA1: CATGTTTGAGCATGGAAACCCGG, chrX:72823838–72823860; gRNA2: GGGTTATGGCGGTCGGTTCCTGG, chrX:73025513–73025535). Subcloning of ES cells was made by limiting dilution. Cells were treated for 24 h with puromycin. As soon as visible, single colonies were picked under a microscope to be screened for deletion by PCR (forward primer: 5′-CGTAGACGCGGCAGTAGTTT-3′, reverse primer: 5′-ACATAAACTCCTTTTCAGGACCA-3′). To identify the targeted allele, we performed a PCR using primers (forward: 5′-CTGTCCAAATGGAGGTGCTT-3′, reverse: 5′-CCTAGGTCCGCTCTCTATCG-3′) that amplify a 203-bp amplicon specifically on the wild-type allele, which contains a single nucleotide polymorphism (SNP; rs29035891). After amplification, PCR products were gel-purified and sequenced using the forward or reverse primer used for PCR. Clones positive carrying the deletion were expanded and differentiated into NPC as previously described[16] and subcloned by limiting dilution. NPC lines were maintained in N2B27 medium supplemented with EGF and FGF (10 ng ml^{-1} each), on 0.1% gelatin-coated flasks. Clones carrying the boundary deletion on the inactive X were identified by RNA FISH against *Xist* with the p510 plasmid probe and DNA FISH with a BAC hybridizing inside the deleted region (RP23-299L1).

RT–PCR and pyrosequencing. RNA extraction and cDNA synthesis were performed with the Cells-to-Ct kit (Ambion) following the manufacturer's instructions. Pyrosequencing primers were designed using the PyroMark Assay Design software. PCR products were purified and annealed with sequencing primers for pyrosequencing using the PyroMark q24 (Qiagen) (*Xist*: forward primer 5′-AGAGAGCC CAAAGGGACAAA-3′, reverse primer 5′-TGTATAGGCTGCTGGCAGTCC-3′, sequencing primer 5′-GCTGGCAGTCCTTGA-3′; *Mecp2*: forward primer 5′-CTGAAGGTTGTAGTGGCTCATG-3′, reverse primer 5′-ATGGTAGCTGG GATGTTAGGG-3′, sequencing primer 5′-CAGAGACAAGCCACTGA-3′; *Huwe1*:forward primer 5′-GCAGGTGTCTGCAAATCCA-3′, reverse primer 5′-GCCGATGTAAAGGCTCCAAG-3′, sequencing primer 5′-TGGGTTCATGT GACAG-3′; *Jarid1c*: forward primer 5′-GCTGCCTCCTTTGCCTGAAAT-3′, reverse primer 5′-TTCAGGGGGCCGCTTACA-3′, sequencing primer 5′-CTCCTTTGCCTGAAAT-3′).

Hi-C read mapping, binning, ICE correction. Hi-C was performed as previously described[27,28]. To obtain allele-specific Hi-C interaction maps in female ES cells ($Xa^{cast}Xa^{129}$) and a derived clonal NPC line ($Xa^{cast}Xi^{129}$) (Methods; Extended Data Fig. 1)[16], we first constructed an allelic genome using the reference mm9 genome and all 19,722,473 SNPs. The allelic (Cast and 129) genomes were then combined to create a reference diploid genome (consisting of 44 chromosomes; 1–19 X,Y,M). All reads were aligned to the diploid genome (as described in ref. 29), thus allowing for a competitive mapping strategy between the two alleles. All reads were trimmed to 50 bp and then aligned using the novoCraft novoalign (v.3.02.00) software package. Reads were aligned using the following options (-r all 5 -R 30 -q 2 -n 50, minimumReadDistance = 5). The best alignment was selected from the list of the top 5 alignments. The alignment was considered unique (allelic), if its alignment score was ≥5 from the second best alignment score (alignment score taken from the ZQ tag). Reads that aligned uniquely to an allele were classified as allelic (either Cast or 129) whereas reads that aligned to both alleles equally (≤5 distance) were classified as ambiguous (AMB) (Extended Data Fig. 1d). Uniquely aligned Hi-C interactions between loci located on the same chromosome were assigned to a specific parental chromosome in *cis* when at least one of the two reads contained a diagnostic SNP, and the other either contained a SNP from the same allele, or mapped to both alleles[30]. We obtained the following paired-end read counts: For ES cells (GUR.2d), a total of 401,684,614 interactions could be aligned combining the two replicates, 372,272,389 of which were unique (after PCR duplicate filter), and 95,650,438 of which could be placed to either the Cast or 129 allele (25.69%). For NPCs (GEI.72b), a total of 277,440,656 interactions could be aligned, 253,254,798 of which were unique (after PCR duplicate filter), and 82,323,031 of which could be placed to either the Cast or 129 allele (32.51%). For ΔFT NPCs (D9B2/B129T3), a total of 229,331,123 interactions could be aligned, 222,941,525 of which were unique (after PCR duplicate filter), and 85,331,870 of which could be placed to either the Cast or 129 allele (38.28%). The difference in percentage of reads assignable to either allele is probably due to differences in the percentage of *cis* interactions found in each sample (biological or technical variation). The 82–95 million read depth supported generation of allele-specific chromatin interaction maps at multiple resolutions (10 Mb, 2.5 Mb, 1 Mb, 500 kb, 250 kb, 100 kb and 40 kb).

Biological replicates were highly correlated. Pearson's correlation coefficients for 500 kb data on chrX were as follows: EHSNP-mF1216__R1R2__chrX-129S1, 0.992331; EHSNP-mF1216__R1R2__chrX-cast, 0.990373; EHSNP-mNPe-deltaRF__R1R2__chrX-129S1, 0.976562; EHSNP-mNPe-deltaRF__R1R2__chrX-cast, 0.983614; EHSNP-mNPe__R1R2__chrX-129S1, 0.990976; EHSNP-mNPe__R1R2__chrX-cast, 0.995202. Autosomes showed similar correlation values. Overall these numbers indicate that the produced Hi-C data was of high quality and well correlated between biological replicates. We pooled all biological replicates into a single Hi-C data set per sample and subsequently used the pooled data for all analyses.

Iterative mapping and error filtering/iterative correction of the chromatin interaction data were performed as previously described[29,31]. Iterative correction was performed on the diploid (44 chromosomes) (replicate pooled) genome-wide matrix for all resolutions.

Hi-C for the TXY (male) samples was performed as previously described[27,28]. Reads originating from the TXY (male) sample was aligned to the mm9 reference genome. Iterative mapping and error filtering/iterative correction of the chromatin interaction data were performed as previously described[29,31]. Iterative correction was performed genome-wide (22 chromosomes) (replicate pooled) on the genome-wide matrix for all resolutions. Biological replicates were highly correlated. We pooled all biological replicates into a single Hi-C data set per sample and subsequently used the pooled data for all analyses.

We obtained the following paired-end read counts: For TXY +dox, a total of 277,191,448 interactions could be aligned, 267,007,192 of which were unique (after PCR duplicate filter). For TXY WT −dox, a total of 308,671,996 interactions could be aligned, 300,102,244 of which were unique (after PCR duplicate filter). For TXY:ΔA +dox, a total of 281,116,218 interactions could be aligned, 273,612,976 of which were unique (after PCR duplicate filter). For TXY:ΔA −dox, a total of 298,436,664 interactions could be aligned, 289,376,893 of which were unique (after PCR duplicate filter).

Allele-specific read mapping validation (Hi-C and ATAC–seq). To validate the accuracy of the allele-specific read alignment strategies used in this paper, we first constructed a set of validation reads tiled across all SNP locations between the Cast and 129 genomes. In brief, for each SNP location on the X chromosome, all overlapping 50-bp reads were extracted (50 total) for each of the 129 and Cast alleles (Extended Data Fig. 1i). All reads were then processed through the Hi-C and ATAC–seq mapping pipelines described in the methods to measure assignment accuracy. Encoded into each fastq readID, was the allelic genome that each read originated from, the relative-position (within the 50-bp read) of the SNP, the base-call of the SNP (A, C, T, G), the chromosome and position (start, end) of the read. The set of validation reads were then processed through the ATAC–seq allele-specific and Hi-C allele-specific pipelines with no modifications. Each processed read was then scored according to whether or not it was correctly placed to not only the correct chromosomal location, but to the correct allele as well. For the Hi-C allele-specific pipeline we found 0 reads assigned to the incorrect chromosomal location and 0 reads assigned to the incorrect allele. All processed validation reads were assigned to the correct coordinate and allele. Reads that span repetitive regions or are of low complexity were inherently filtered via the Hi-C pipeline and thus would be excluded from both the actual data and from the set of validation reads. For the ATAC–seq allele-specific pipeline, we found that for reads coming from the 129 X chromosome, only 0.09% mapped to the Cast chromosome (only 3 of these fell within ATAC–seq peaks). For reads coming from the Cast X chromosome, only 0.21% mapped to the 129 chromosome (5 of which fell within ATAC–seq peaks). Furthermore, we feel that our analysis may be even more accurate when using paired-end reads as we do for all ATAC–seq data analysis (Extended Data Fig. 1i).

Generation of Xist-positive Hi-C signal for comparison with DNA FISH. *Xist* RNA FISH performed in parallel with Hi-C on the same inducible *Xist* ES cell samples revealed that *Xist* expression (either wild-type or A-repeat mutant) was induced in ~35–45% of dox-treated cells. Hence the Hi-C signal can be represented as the sum of ~35–45% of reads coming from Xist-positive X chromosomes,

and ~55–65% of reads that are generated from non-coated X chromosomes. For example,

$$\text{Hi-C}(+\text{dox}) = 35\% \times \text{Hi-C}(\text{Xist}+) + 65\% \times \text{Hi-C}(\text{Xist}-)$$

We took advantage of the fact that the signal from Xist− cells is measured independently in the non-induced (−dox) sample:

$$\text{Hi-C}(\text{Xist}-) = \text{Hi-C}(-\text{dox})$$

to extract the signal of Xist-positive cells:

$$\text{Hi-C}(+\text{dox}) = 35\% \times \text{Hi-C}(\text{Xist}+) + 65\% \times \text{Hi-}C(-\text{dox})$$

and hence

$$\text{Hi-C}(\text{Xist}+) = [\text{Hi-C}(+\text{dox}) - 65\% \times \text{Hi-C}(-\text{dox})] \times (1/0.35).$$

Negative values were assigned to 'not analysed' (NAs). The Hi-C(Xist+) signal was then used to compare Hi-C data with RNA/DNA FISH experiments in Fig. 4b, which allow to discriminate *Xist*-coated and non-coated chromosomes visually by the presence of an *Xist* cloud. In Fig. 4c, *Xist*− signals correspond to the −dox sample.

Hi-C SNP density filter. To remove potential biases in the Hi-C data related to the density of SNPs in each bin, we calculated the number of SNPs residing in each genomic interval (bin) for all Hi-C bins across all bin sizes. We then calculated the median number of SNPs per bin, and produced a minimum required SNP density cutoff defined as the (median − 1.5 × IQR). Any bins with less SNPs than the cutoff were removed from all analyses. The SNP density cutoffs used for each bin size were: 40 kb, 43 SNPs; 100 kb, 216 SNPs; 250 kb, 776.5 SNPs; 500 kb, 1,767.25 SNPs. The non-SNP-density-filtered data was only used for visualization purposes (figure heat maps). Throughout the manuscript, we refer to Hi-C as data that has been iteratively corrected[31] and run through the SNP-density filter.

Compartment analysis. The presence and location of the A/B compartments were calculated as previously described[32]. Compartments were derived from the 250-kb iteratively corrected Hi-C data for each chromosome separately using the CIS maps for each sample/allele. The code used to generate the compartments (PC1 from PCA analysis) is publicly available on Github (matrix2compartment.pl): https://github.com/dekkerlab/giorgetti-nature-2016. Compartments were generated all default options except the (cis alpha) option, set to (-ca 0.005).

Insulation and boundary calculation. TAD structure (insulation/boundaries) was defined via the insulation method as previously described with minor modifications[32]. The code used to calculate the insulation score is publicly available on Github (matrix2insulation.pl): https://github.com/dekkerlab/giorgetti-nature-2016. Insulation vectors were detected using the following options: (-is 480000 -ids 320000 -im iqrMean -nt 0 -ss 160000 -yb 1.5 -nt 0 -bmoe 0). The output of the insulation script is a vector of insulation scores, and a list of minima along the insulation vector (inferred as TAD boundaries). The TAD boundaries were not used in this study.

Interaction pile-up maps. Interaction pile-up maps were constructed from all pairwise interactions between either the list of 87 wild-type NPC Xi chromosome escapees or the 29 ΔFT NPC Xi chromosome escapees. Using the 40 kb Hi-C data, a 2-Mb window centred around each pairwise interaction (pixel) was taken (25 bins in each direction, yielding 51 × 51 sub-matrix). Any resulting sub-megabase that overlapped the ($y = x$) diagonal in the matrix was excluded from the analysis (effectively excluding all interactions <2 Mb). All sub-matrices were then averaged to produce the final (mean) pile-up map. A strong signal at the centre suggests that the elements used tend to contact one another in 3D space.

The Xi chromosome is as accessible and detectable in Hi-C as the Xa chromosome and autosomes. The number of RAW reads observed for both the Xa and Xi chromosomes were very similar for all chromosomes, thus demonstrating that the Xi chromosome is not simply less accessible/visible to the Hi-C methodology. ES-cell-chrX-129S1, 1,118,327; ES-cell-chrX-Cast, 1,104,709; NPC-chrX-129S1, 1,147,072; NPC-chrX-Cast, 1,148,128; ΔFTNPC-chrX-129S1, 1,314,476; ΔFTNPC-chrX-Cast, 1,288,802. Bias in read directional due to partial digestion is typically observed up to ~10 kb. For interactions between fragments separated by over 10 kb this bias is negligible, indicating at least one digestion occurring between them in every cell. This genomic distance is therefore a measure for digestion efficiency[27]. For both the Xa and Xi chromosomes, this genomic distance is ~6–10 kb, indicating that digestion efficiency of chromatin on the Xa and Xi chromosomes are comparable. Thus, the unique conformation of the Xi chromosome does not affect Hi-C analysis, as was also found for condensed mitotic chromosomes[28].

Correlation analysis of Hi-C matrices. We compared X chromosome interaction matrices of *Xist*-inducible lines (pre/post-*Xist* induction in wild-type and A-repeat mutant samples) with that of the NPC Xi chromosome, at 500-kb resolution. As distance-dependent decay of interaction frequency causes all interaction matrices to be highly correlated, we first compensated for this effect by multiplying the read count in each bin by its respective genomic distance. We then calculated the Spearman correlation between each pair of matrices. After wild-type *Xist* induction in TXY cells, the interaction map becomes more similar to that of the NPC Xi chromosome (rho = 0.17 to 0.31), while no change is observed in the repeat-A mutant (rho = 0.17 to 0.17).

RNA and 3D-DNA FISH. FISH was performed as previously described[33]. ES cells and NPCs were cultured on gelatin-coated coverslips #1.5 (1 mm) and fixed in 3% paraformaldehyde for 10 min at room temperature. Cells were permeabilized on ice for 5 min in 1× PBS, 0.5% Triton X-100 and 2 mM vanadyl-ribonucleoside complex (VRC, New England Biolabs), and coverslips were stored in 70% ethanol at −20 °C. Before FISH, samples were dehydrated through an ethanol series (80%, 95%, 100% twice) and air-dried briefly. For RNA FISH, cells were directly hybridized with denatured probes. For DNA FISH, samples were first denatured in 50% formamide/2× SSC (pH = 7.3) at 80 °C for 37 (ES cell) and 35 (NPC) min, immediately placed on ice and washed twice with ice-cold 2× SSC. After overnight hybridization at 37 °C for RNA FISH or 42 °C for DNA FISH, coverslips were washed at 42 °C for RNA or 45 °C for DNA, three times for 5 min in 50% formamide/2× SSC at pH 7.3, and three times for 5 min in 2× SSC. Nuclei were counterstained with 0.2 mg ml^{-1} DAPI (2 mg ml^{-1} for structured illumination microscopy), further washed twice for 5 min in 2× SSC at room temperature and finally mounted with 90% glycerol, 0.1× PBS, 0.1% *p*-phenylenediamine at pH9 (Sigma).

RNA FISH probes. We used the p510 plasmid coupled with Cy5 to detect *Xist*. For RNA FISH on escape genes, we used the following BAC and fosmid probes: RP23-436K3, RP23-328M22, RP24-436K3, WI1-1269O10 (*Mecp2*), RP24-157H12 (*Huwe1*), RP23-13D21 (*G6pdx*), RP24-148H21 (*Jarid1c*).

DNA FISH probes. In experiments to detect the mega-domain boundary, fluorescent oligonucleotides (average length 45 bp, 5′-modified with Atto 448 or Atto 550, average density: one oligonucleotide every 3 kb) were obtained from MYcroarray Inc. Oligonucleotides were designed to tile the following consecutive 18-Mb regions: chrX:35,000,000–53,000,000, chrX:53,000,000– 72,000,000, and chrX:72,000,000–90,000,000. To detect the *DXZ4* region we used the RP23-299L1 BAC.

Imaging and quantification of 3D DNA FISH. Three-dimensional image stacks (200 nm distance between consecutive *xy* planes) were acquired on a DeltaVision Core wide-field microscope (Applied Precision) equipped with a CoolSNAP HQ2 camera operated at 2X binning, and a 100× PlanApo oil immersion objective (the effective pixel size was 129 × 129 nm). Xi chromosome signals were identified via the presence of an *Xist* mRNA cloud in the far-red channel (p510-Cy5 probe). Pearson correlation between red and green signals was calculated using custom-made ImageJ macros as follows. After subtracting the background from each *xy* plane (generated by morpholoigcal opening the image with a circle of 5 pixels in radius), Pearson correlation between red and green pixel intensities was measured inside a fixed-size region of 40 × 40 × 20 pixels (5.16 × 5.16 × 4 μm^3) centred on each FISH signal. The significance of Xi versus Xa chromosome differences in correlation was assessed by Wilcoxon's rank sum test. Random nuclear positions were used to estimate the background correlation that could be observed due to non-specific probe hybridization.

The gyration tensor of a greyscale image is defined as

$$S_{ab} = \sum_k I_k(r_a^k - r_a^{CM})(r_b^k - r_b^{CM}) / \sum_k I_k,$$

where k is an index running over voxels, I_k is the greyscale intensity of voxel k, and r_a^k and r_a^{CM} are the a-th components (x, y, or z) of the xyz position of voxel k, and of the centre of mass of the image, respectively. The gyration tensor was valuated in a region of interest of 3.8 × 3.8 × 4 μm^3 centred on each FISH signal and the gyration radius was calculated as

$$R_g = \sqrt{\lambda_1 + \lambda_2 + \lambda_3}$$

where $\lambda_{1,2,3}$ are the eignevalues of S_{ab}.

RNA-seq. RNA-seq data for the ES cell (GUR.2d) and NPC (GEI.72b) was obtained from previously published work (PMID 24576422)[16]. RNA-seq data for the mutant NPC (D9B2/B129T3) was obtained and processed as previously described[16].

RNA-seq 'expressed/escapee' classification. The allelic RPKM values were derived for each gene by splitting the RPKM value by the 129 ratio. 129 RPKM = (RPKM × 129 ratio); Cast RPKM = (RPKM × (1 − 129 ratio)). Any gene with an allelic RPKM value ≥3 RPMK was classified as being expressed. Any gene expressed on the Xi chromosome was classified as being an escapee.

ATAC–seq. ATAC–seq library preparation was performed exactly as previously described[34]. Sequencing was carried out on an Illumina NextSeq 500 generating 2 × 75 bp paired-end reads. Libraries were sequenced to a depth of 25–35 million reads per sample. Reads were trimmed using CutAdapt and aligned using Bowtie2.

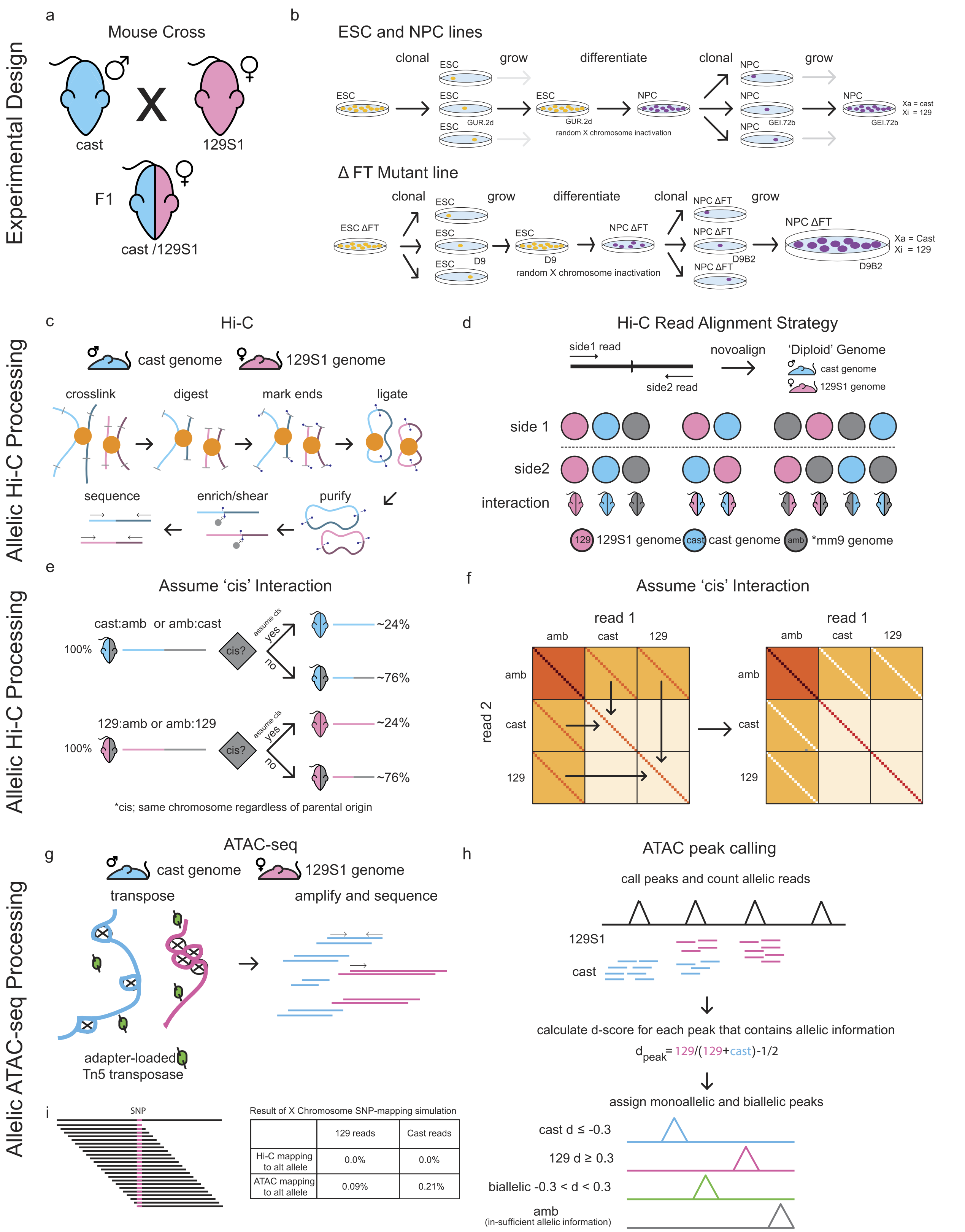
Reads were aligned to a custom 129/CastEiJ genome in which SNP sites were replaced by 'N'. Approximately 52–58% of reads per line contained 'N's and were assigned to the 129 or Cast allele based on the identity of the base at that location. Reads containing non-concordant SNPs were rare and were discarded. Reads not containing SNP sites were included in overall peaks but not were excluded from allele-specific tracks. ATAC–seq peaks were called using MACS2 with no shifting model. For TXY *Xist*-inducible male cell lines, which contain only one X chromosome, peaks were called after normalizing all samples by read count on the autosomes. The set of X chromosome peaks was defined as the merge of peaks from all samples, and bedtools was used to calculate coverage within peaks.

Assigning allele-specific ATAC–seq peaks. For each ATAC–seq peak, all N-containing reads were counted and assigned to 129 or Cast alleles based on SNP at the N-containing position. To assign monoallelic and biallelically accessible peaks, a d-score was calculated as a measure of allelic imbalance[35]. In brief, for a given peak the d-score was calculated as the ratio of 129 reads to total number of reads $-$ 1/2. A peak with a d-score ≥ 0.3 was assigned as a 129-specific peak. A peak with a d-score ≤ -0.3 was assigned as a Cast-specific peak. Any peak with a d-score > -0.3 was assigned as a peak in 129 (monoallelic or biallelic). Any peak with a d-score $< +0.3$ was assigned as a peak in Cast (monoallelic or biallelic).

Annotating ATAC–seq peaks using ChIP–seq data. ATAC–seq peaks were annotated using existing published ChIP–seq data sets. CTCF ChIP–seq came from whole female mouse brain[36]. Called CTCF binding sites were used and extended $\pm$300 bp before overlapping with ATAC–seq peaks. H3K27ac and p300 ChIP–seq are from mouse NPCs[37]. For H3K27ac and p300 ChIP–seq data, peaks were called using MACS2 and then overlapped with ATAC–seq peak locations.

Integrating Hi-C, ATAC–seq and RNA-seq data. Integrative analysis of Hi-C insulation (TAD structure), ATAC–seq counts, and RNA-seq RPKM was performed as follows. A promoter region was defined for each gene as $\pm$500 bp from the TSS. ATAC peaks were assigned to a gene if they overlapped with the promoter region. In the event that $>$1 ATAC peak overlapped with the promoter, the closer ATAC peak was chosen. An ATAC count of 0 was assigned to each promoter, if it did not contain an ATAC peak. If the ATAC allelic counts overlapping the promoter were $<$10, then the ATAC count was set to 'NA'. The 40-kb bin overlapping the promoter region was used to display the insulation and insulation-difference value.

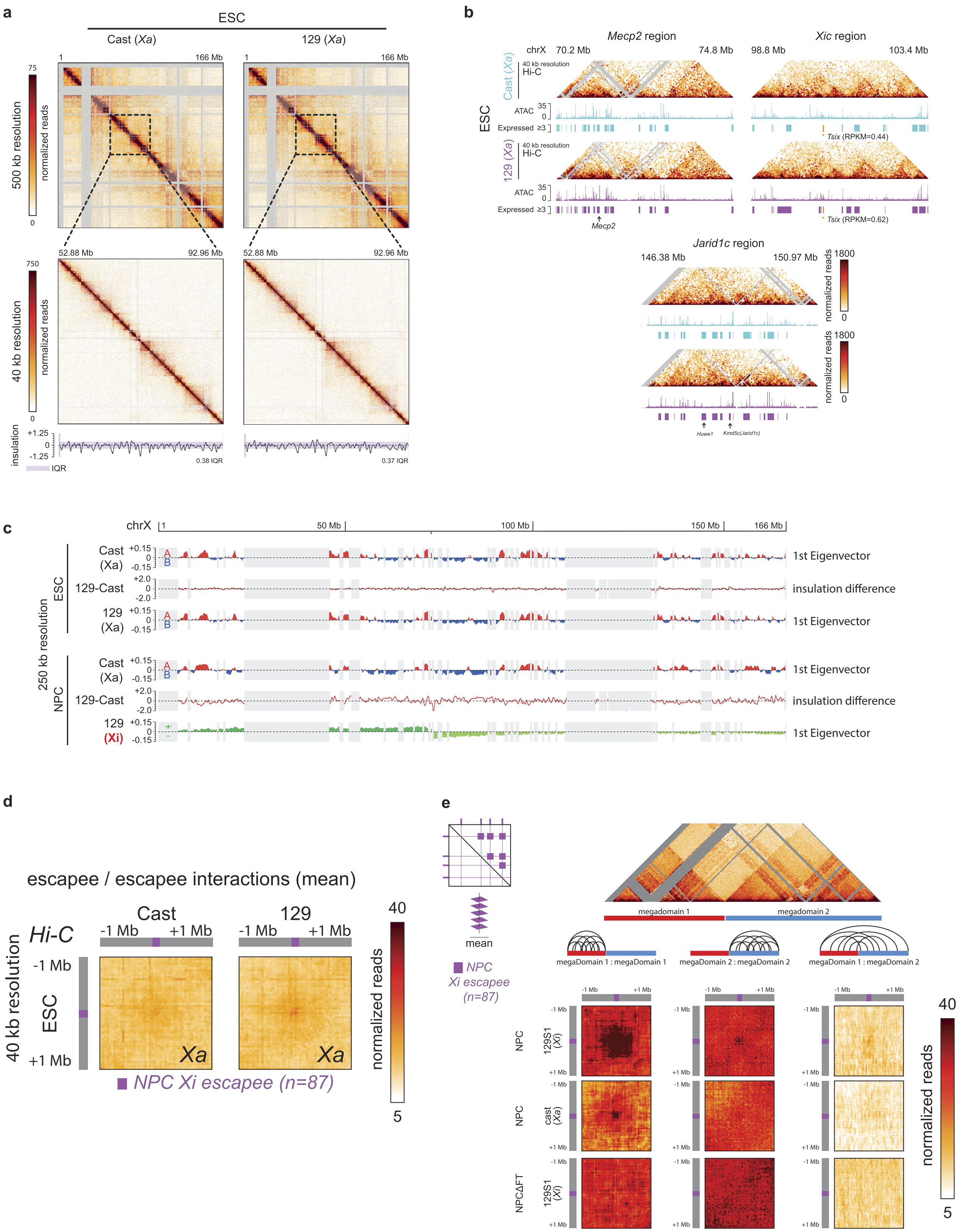
27. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
28. Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342,** 948–953 (2013).
29. Lajoie, B. R., Dekker, J. & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* **72,** 65–75 (2015).
30. Selvaraj, S., R Dixon, J., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.* **31,** 1111–1118 (2013).
31. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9,** 999–1003 (2012).
32. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523,** 240–244 (2015).
33. Chaumeil, J., Augui, S., Chow, J. C. & Heard, E. Combined immunofluorescence, RNA fluorescent *in situ* hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation. *The Nucleus* **463,** 297–308 (2008).
34. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
35. Eckersley-Maslin, M. A. *et al.* Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev. Cell* **28,** 351–365 (2014).
36. Berletch, J. B. *et al.* Escape from X inactivation varies in mouse tissues. *PLoS Genet.* **11,** e1005079 (2015).
37. Martynoga, B. *et al.* Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev.* **27,** 1769–1786 (2013).
38. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521,** 232–236 (2015).



	129 reads	Cast reads
Hi-C mapping to alt allele	0.0%	0.0%
ATAC mapping to alt allele	0.09%	0.21%

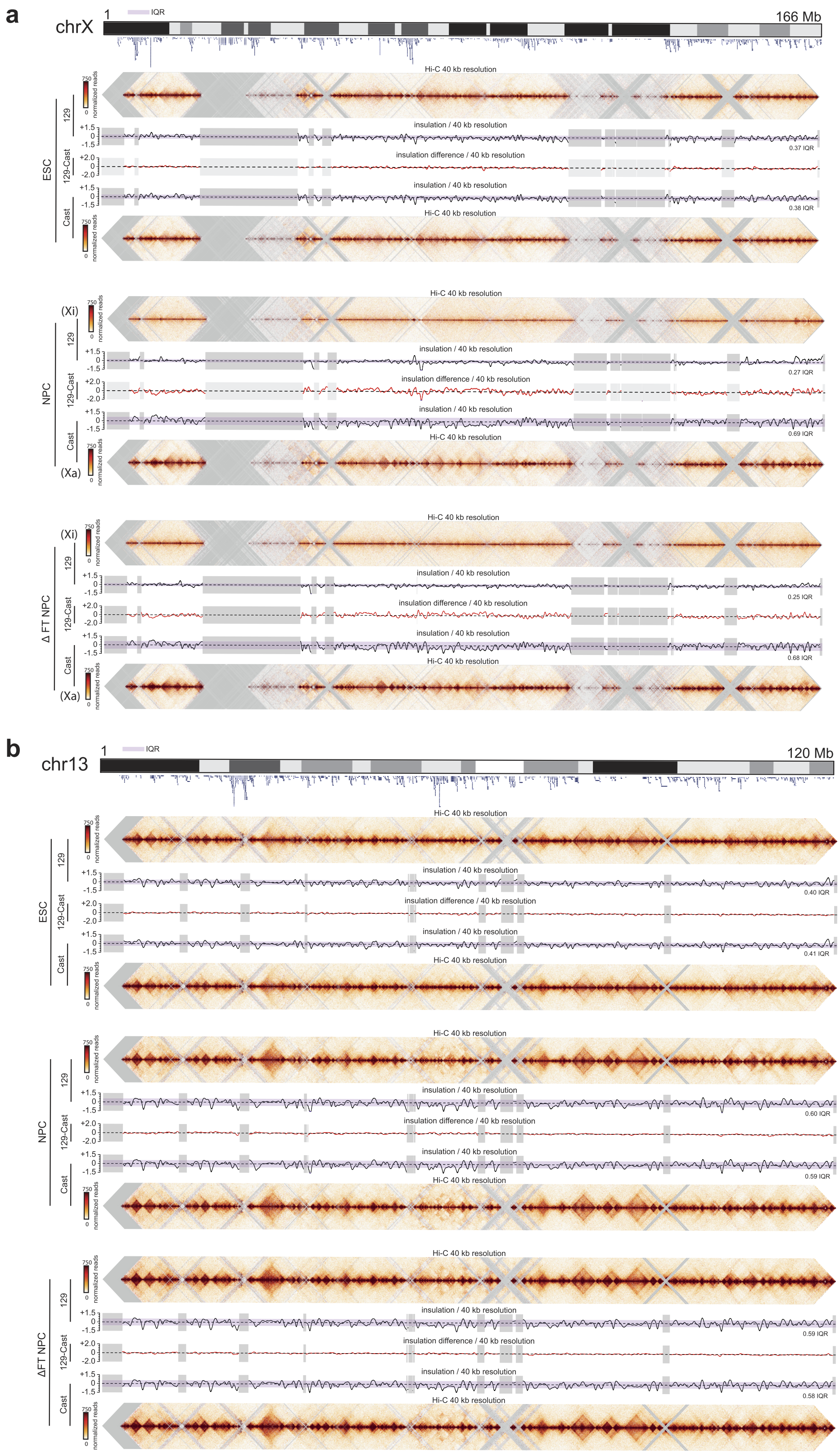
Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Description of allele-specific Hi-C and ATAC–seq. a, Schematic of hybrid mouse strains used for all experiments. **b**, Top, scheme outlining differentiation of ES cells to NPCs and picking of clones. Bottom, scheme outlining CRISPR deletion of the mega-domain boundary in ES cells, differentiation to NPCs and the picking of clones. **c**, Schematic of Hi-C library generation. **d**, Schematic of the Hi-C alignment strategy. Paired-end reads are aligned to a 'diploid' genome consisting of 22 chromosomes from Cast, and 22 chromosomes from 129 (1–19 X, Y, M). The interaction row shows all possible paired-end read combinations between the 129, Cast and ambiguous (AMB) genomes. **e**, Schematic showing the re-assignment of a subset of '*cis*' interactions. Paired-end reads in which one side uniquely aligned to an allele and the other side aligned equally to both alleles (AMB), were re-classified as allelic reads, only if both reads aligned to the same chromosome (*cis*). **f**, Cartoon explaining the re-assignment of 129:amb or cast:amb *cis* interactions. **g**, Scheme for ATAC–seq library preparation. Cells are lysed followed by incubation with adaptor-loaded hyperactive Tn5 transposase. The transposase integrates into accessible DNA, and these fragments are then directly amplified and sequenced. **h**, Scheme for allele-specific ATAC–seq data analysis. **i**, SNP-mapping simulation. For each SNP location on the X chromosome, all overlapping 50 bp reads were extracted (50 total) for each of the 129 and Cast alleles. All reads were then processed through the Hi-C and ATAC–seq mapping pipelines described in the Methods to measure assignment accuracy. Results are shown in the table.

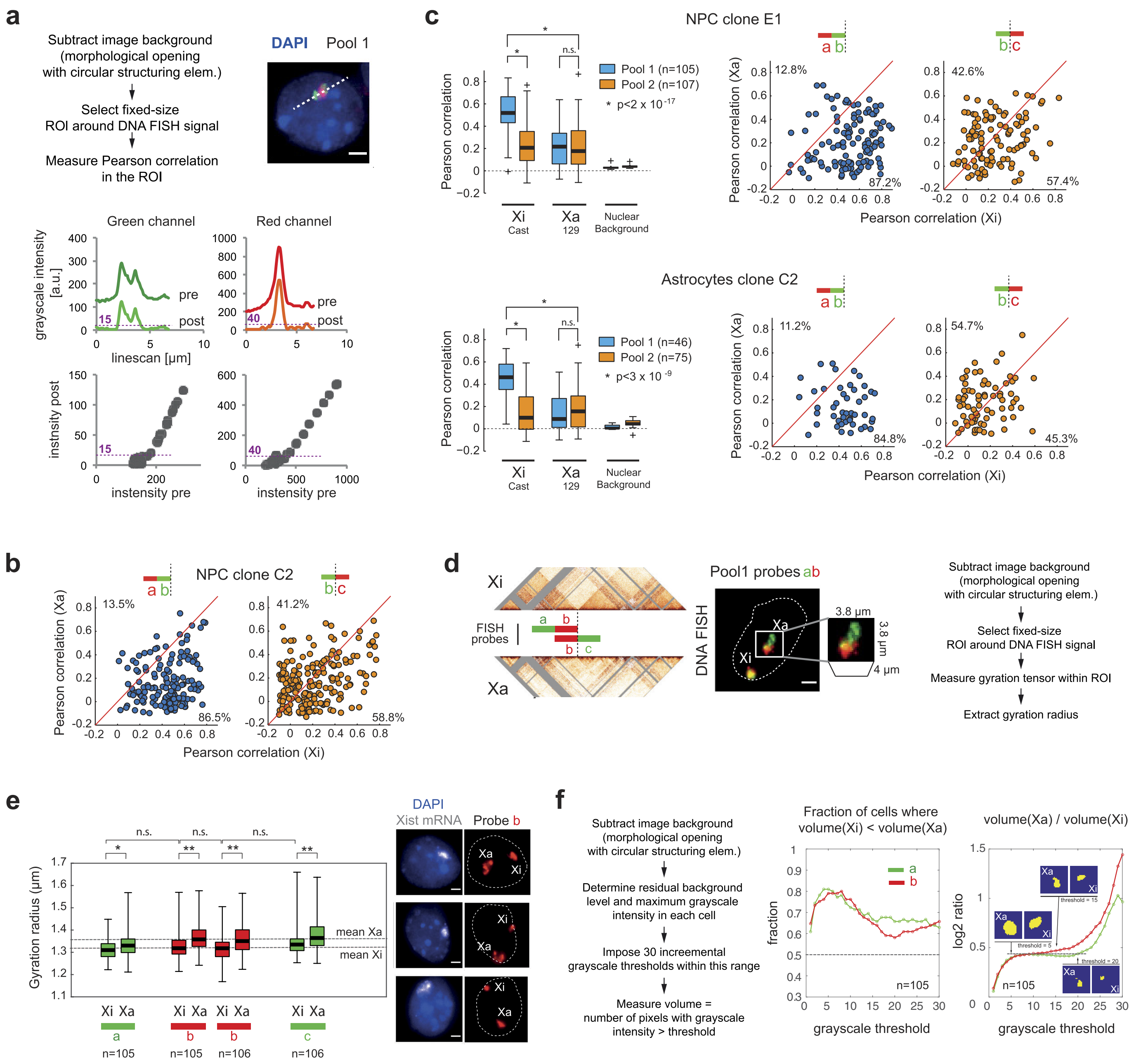


Extended Data Figure 2 | See next page for caption.

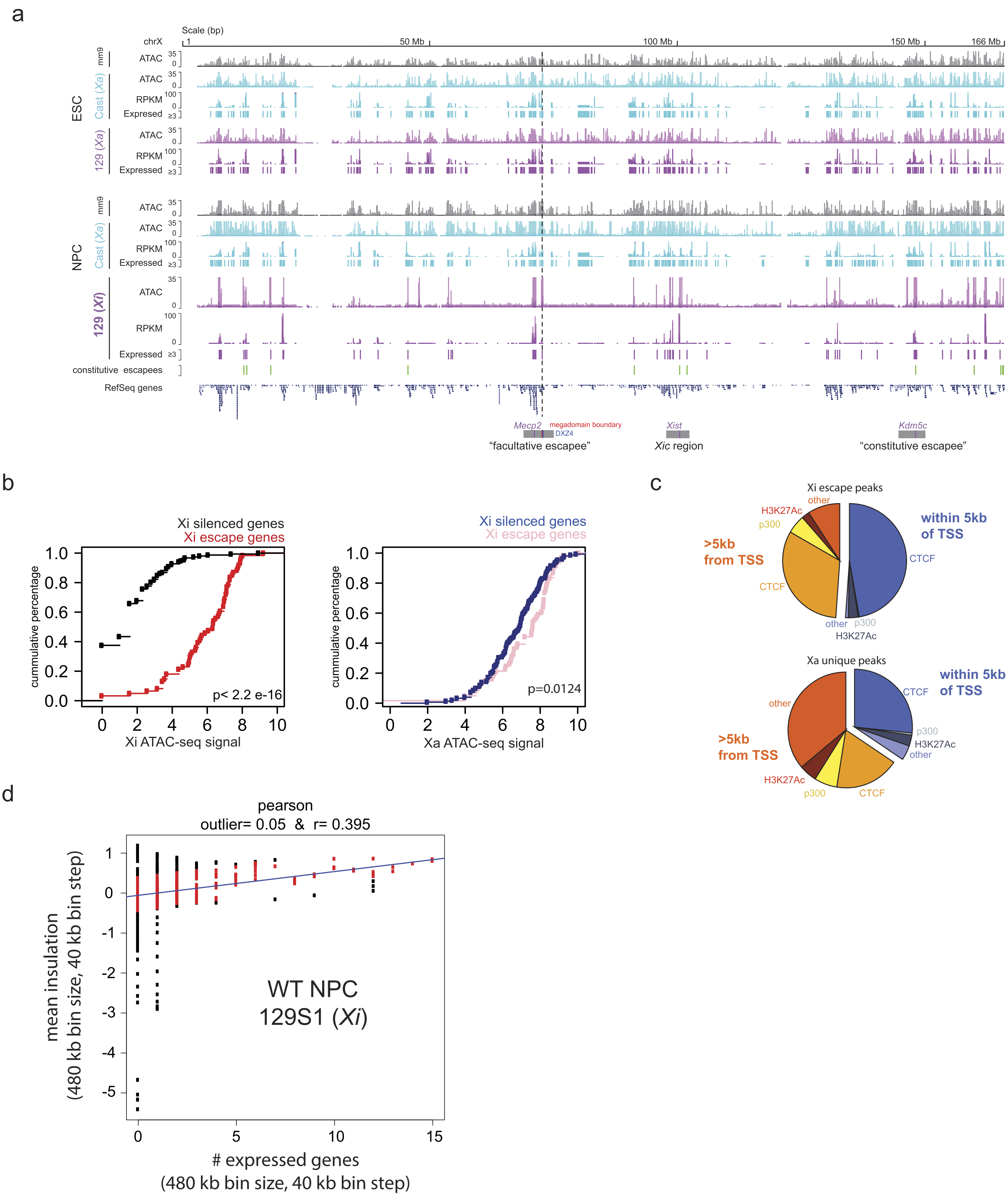
Extended Data Figure 2 | Structure of the Xa chromosome in ES cells, compartment analysis and characterization of interactions between genes that escape XCI. a, Allele-specific Hi-C contact maps for X chromosome in ES cells at 500-kb resolution (top), and for a ~40-Mb region centred around the *DXZ4*-containing locus at 40-kb resolution (bottom). The insulation score is plotted at the bottom of each 40-kb heat map as in Fig. 1a. Purple shaded areas indicate the IQR of insulation scores along the chromosome. **b**, Zoomed in view of three regions showing Hi-C interactions, RNA-seq and ATAC–seq signal on the Cast and 129 Xa chromosomes in ES cells. **c**, Compartment profiles of X chromosome in ES cells and NPCs. The first eigenvector (PC1) of each allele-specific Hi-C contact map, obtained with principal component analysis, is shown, together with the difference in chromosome-wide insulation score between the 129 and Cast allele. A/B-compartments are evident in ES cells and NPCs along both Xa (red and blue signal), whereas the first eigenvector corresponds to the two mega-domains for the Xi chromosome in NPCs. In ES cells, both Xa chromosomes display comparable insulation profiles (difference is close to 0 along the chromosome), whereas in NPCs large differences are observed (difference in insulation fluctuates along the chromosome). Grey areas indicate regions with low SNP density that were excluded from analysis. **d**, Same plots as in Fig. 2e for the same set of genes that escape XCI in NPCs, but using Hi-C data obtained in ES cells. **e**, As in **d**, where the interactions between loci that escape XCI on the wild-type Xi chromosome were divided in three different groups, according to whether pairs of loci belong to the same or different mega-domains. Interactions across the mega-domain boundary only occur on the wild-type Xi chromosome and are lost on the ΔFT Xi chromosome concomitant with loss of transcription of constitutive escapees.



Extended Data Figure 3 | Comparison of allele-specific Hi-C maps of chromosomes X and 13 in ES cells, NPCs and boundary-mutant ΔFT NPCs. a, Hi-C data, insulation scores, and the difference in insulation scores (129 Cast) are shown for ES cells, NPCs and mutant ΔFT NPCs for both alleles (Cast and 129) for the X chromosome. Large dips in the insulation vector are found at TAD boundaries. Peaks in the insulation vector are found towards the centre of each TAD. The insulation difference plot highlights areas of differential TAD structure between the alleles (many differences along the X chromosome in NPCs as compared to the allelic differences along autosomes, see **b**). **b**, Same as **a** for chromosome 13. The insulation difference plot highlights areas of differential TAD structure between the alleles (rare).

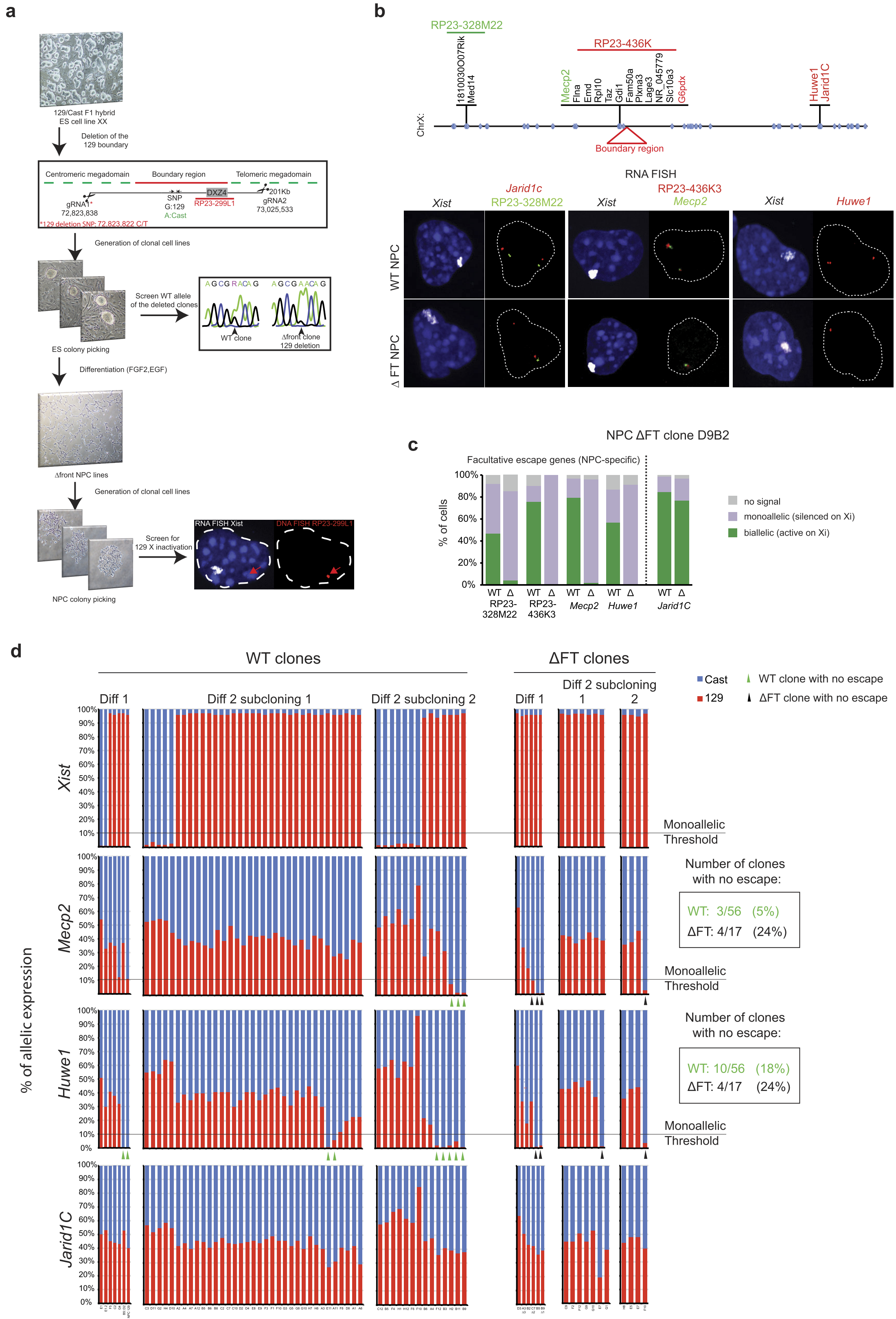


Extended Data Figure 4 | Quantitative analysis of 3D DNA FISH on the Xa and Xi chromosomes. a, Top left, scheme of the procedure used to quantify the Pearson correlation. A background is generated for each *xy* plane in a three-dimensional *z*-stack by morphological opening the image with a circle of 5 pixels in radius, and subtracted from it. Pearson correlation between red and green pixel intensities is measured inside a fixed-size region of 40 × 40 × 20 pixels (5.16 × 5.16 × 4 µm) centred on each FISH signal. To demonstrate that background subtraction does not affect the measured correlations, we show here a line-scan of 10 µm across a typical DNA FISH signal (top right). The shape of the signals along the line scan, as well as their relative intensities, is not affected by background subtraction (bottom). Pre- and post- refer to line scans taken on images before and after background subtraction, respectively. **b**, In more than 80% of nuclei in NPCs, Pearson correlations are higher on the Xi than on the Xa chromosome. Shown is NPC clone C2 (the same where Hi-C was performed). **c**, Same quantification as in Fig. 1c (and **b**) for an independent NPC clone (E1) where the active X is on the 129 allele and the inactive X chromosome on the Cast, and in astrocytes derived from NPC clone C2. *n* denotes the number of cells analysed in DNA FISH. Centre lines represent medians; crosses denote experimental points that were considered as outliers (larger than (q3 + 1.5 × (q3 − q1) or smaller than q1 − w(q3 − q1), where q1 and q3 are the twenty-fifth and seventy-fifth percentiles of the data distribution, respectively). **d**, Scheme of the gyration tensor based analysis of FISH volumes (see Methods). **e**, Left, gyration radii of DNA FISH signals from probes a, b and c. Probe b was used in combination with both probes a and c separately in two independent experiments. Statistical significance was assessed by Wilcoxon's rank sum test (*$P < 0.05$, **$P < 1 \times 10^{-5}$). The mean gyration radii for Xa and Xi chromosome signals are indicated by dotted lines as a guide for the eye. On the Xi chromosome, the mean gyration radius is approximately 6% smaller than on the Xa chromosome corresponding to approximately 20% smaller volume. Right, representative images of probe a, showing smaller size and increased roundness of the Xi chromosome signals. **f**, Left, scheme of the thresholding-based method for volume quantification. Thirty increasing threshold levels were imposed, starting from the residual grayscale background level surrounding the signal, up to the minimum between the red and green channel grayscale maxima. For each of these thresholds we determined the number of voxels in each channel, where the greyscale intensity was higher than the threshold. Centre, the fraction of cells where the Xa chromosome signal is larger than the Xi chromosome is between 60% and 80% in the entire threshold range. Right, in a wide range of thresholds, the volume of Xa chromosome signals is approximately 25% bigger than Xi chromosome signals. Results are shown here for probes a and b; the same holds for probe c (not shown). *n* denotes the number of cells analysed in DNA FISH. Centre lines: medians. Boxes: middle 50% of data points. Two biological replicates were analysed for each experiment.

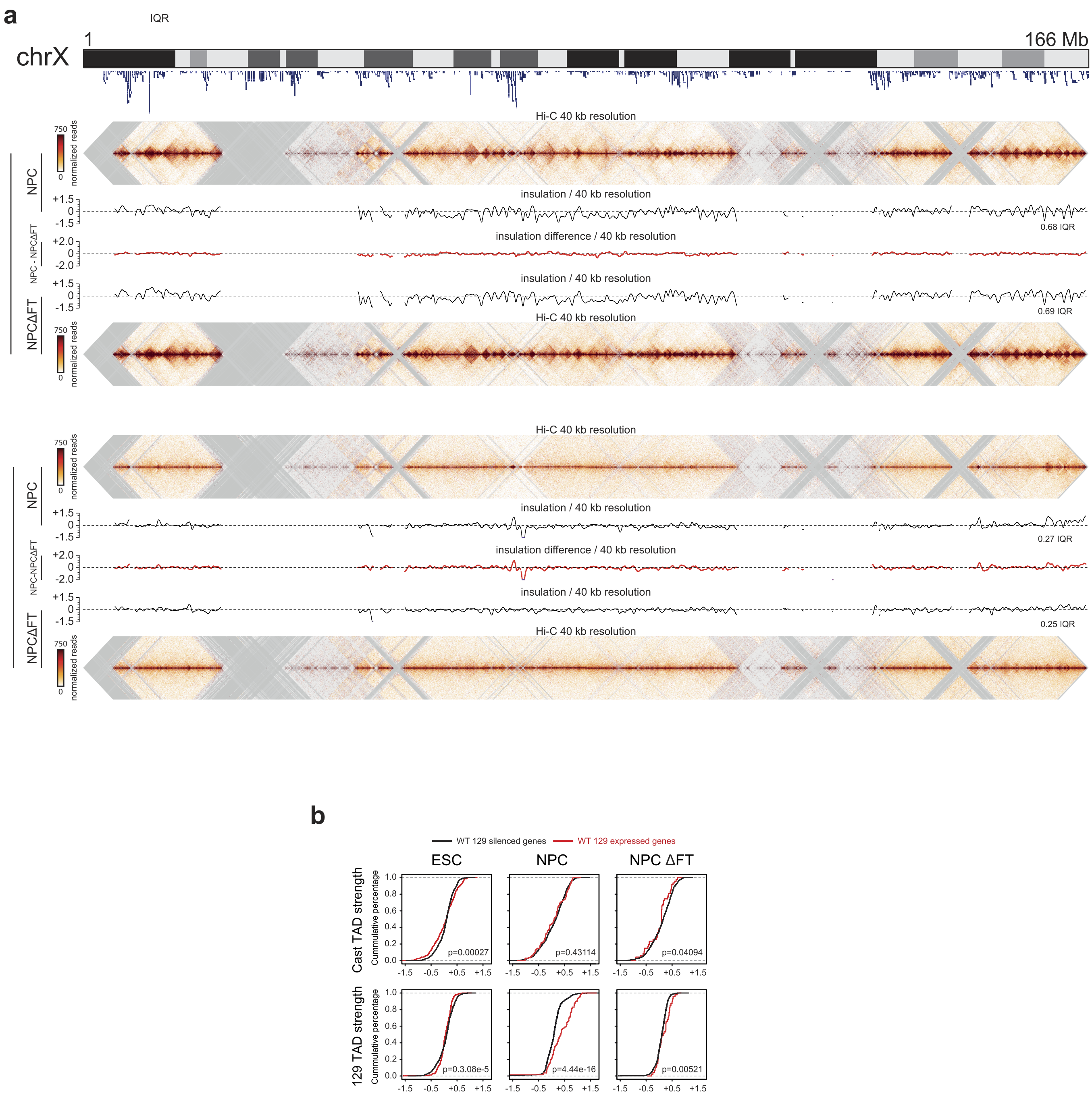


Extended Data Figure 5 | See next page for caption.

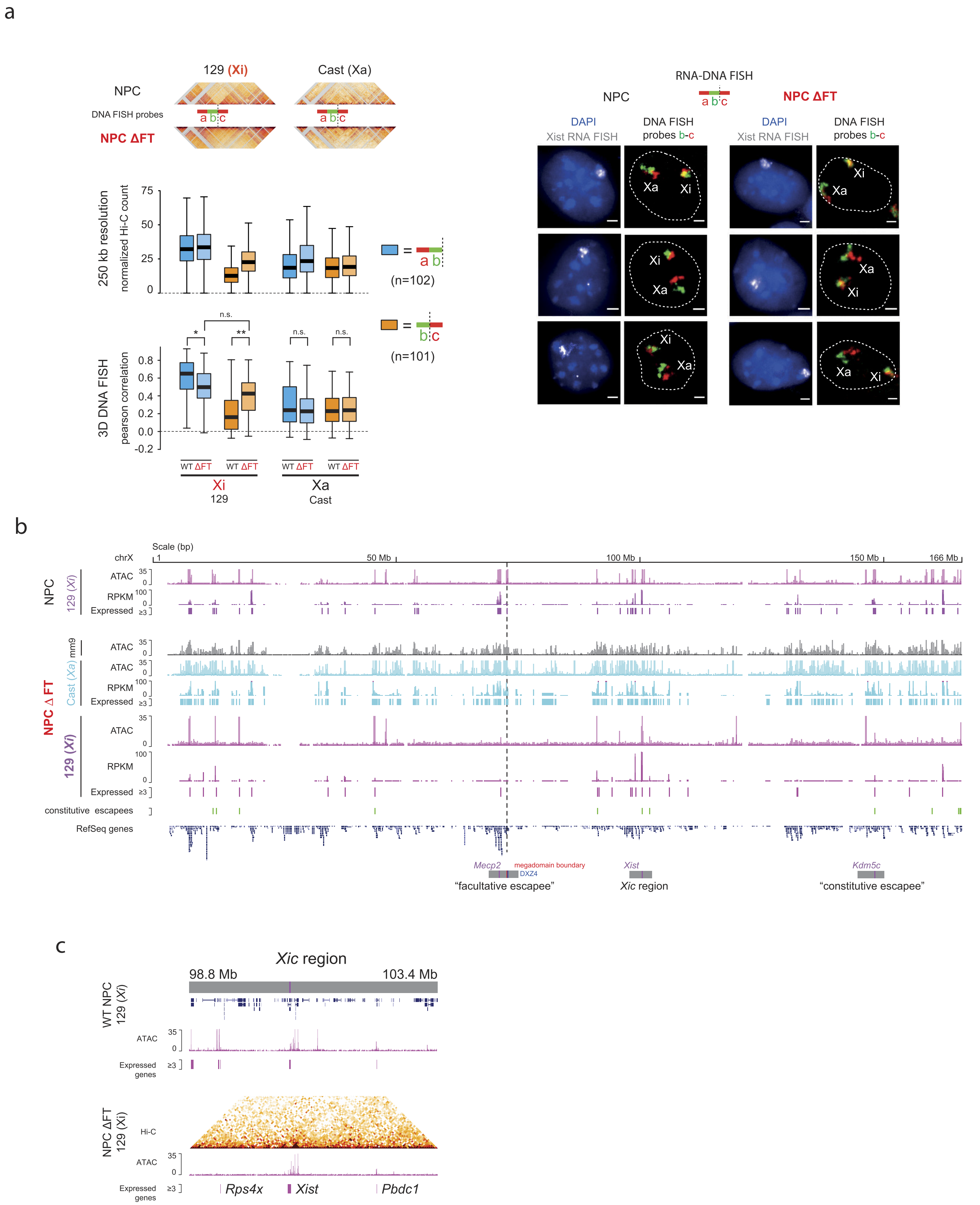
Extended Data Figure 5 | Integrative analysis of allele-specific Hi-C, RNA-seq and ATAC–seq. a, X-chromosome-wide ATAC–seq and RNA-seq in ES cells and NPCs. ATAC shows signal for ambiguous, 129- and Cast-specific reads in ES cells and NPCs. RNA-seq shows total signal as well as expressed gene calls. ATAC–seq shows global loss of chromatin accessibility and expression on the Xi chromosome, except at specific locations that mostly overlap with escape genes. Dotted line denotes mega-domain boundary. Position of constitutive escapees was adapted from ref. 3. **b**, Escape genes on the Xi chromosome (as determined by RNA-seq) fall within regions with high ATAC–seq signal (Kolmogorov–Smirnov test $P < 2.2 \times 10^{-16}$). **c**, Pie charts showing the distribution of peaks that escape XCI versus the peaks that are unique to the Xa chromosome. Peaks are classified into those that are promoter-proximal (within 5 kb of TSS) and distal (>5 kb from TSS). Annotations are based on binding sites identified by ChIP–seq[37,38]. **d**, Scatter plot showing the relationship between TAD signal (insulation score) and number of expressed genes of the wild-type NPC 129 (Xi) Hi-C data set. The *y* axis shows the mean insulation score in a 480-kb window with a 40-kb step size. The *x* axis shows the number of expressed genes in a 480-kb window with a 40-kb step size. The Pearson R value (0.395) is shown above and is calculated on only the middle 90% (red points) of the residuals. Outliers are shown in black (bottom 5% and top 5% of residuals)



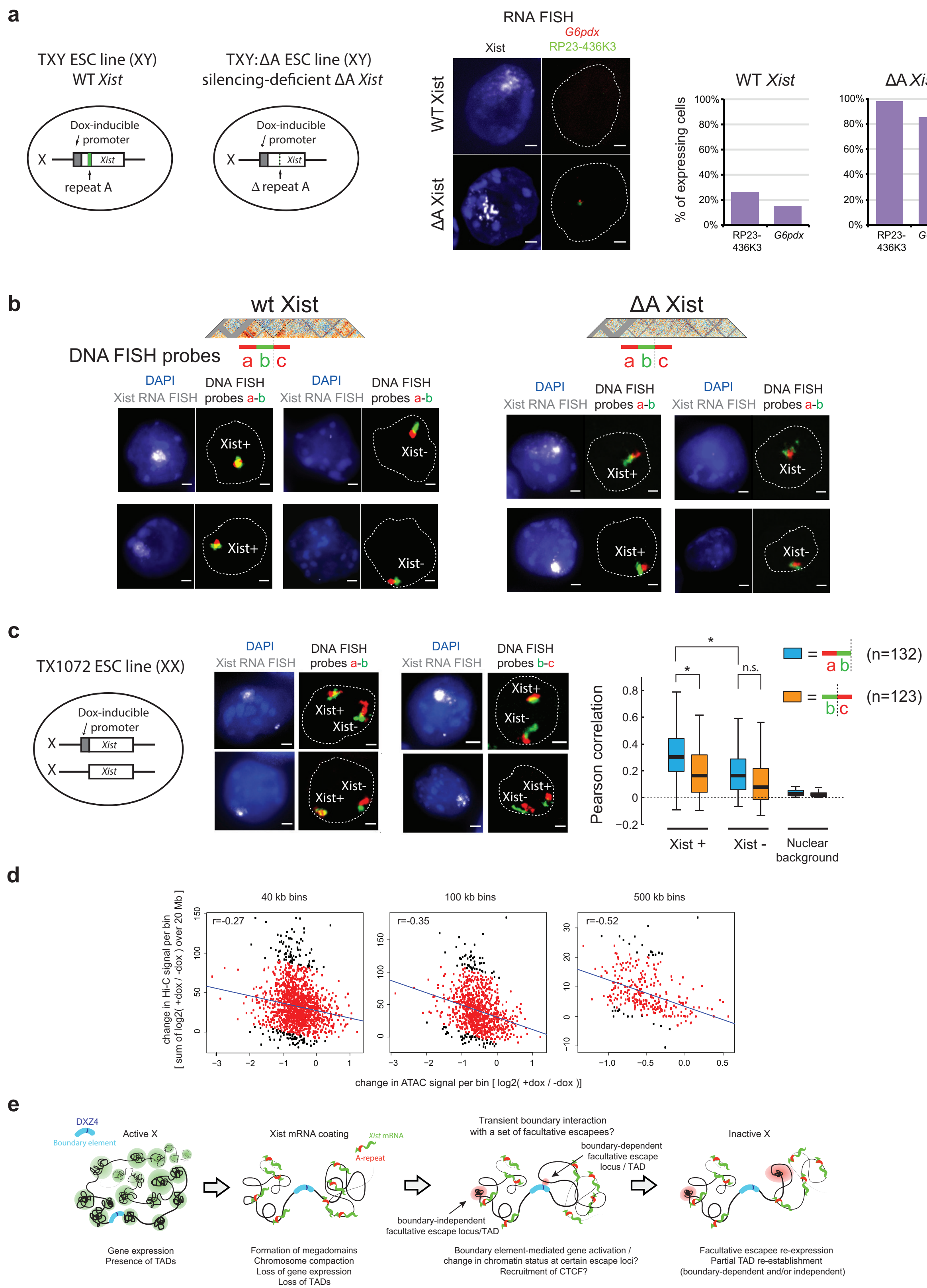
Extended Data Figure 6 | Analysis of wild-type and boundary-mutant ΔFT NPCs by FISH and RT–PCR followed by pyrosequencing. **a**, Scheme of the strategy used to delete the mega-domain boundary region in ES cells and to derive ΔFT NPCs. **b**, RNA FISH against constitutive and facultative escapees confirms RNA-seq and ATAC–seq results in the ΔFT NPC clone D9B2. Top, the positions of BAC probes (RP23-328M22 and RP23-436K) are shown relative to the escape genes that they span. Coloured gene names correspond to transcripts that were detected with specific fosmid probes. Bottom, sample RNA FISH images showing that expression of facultative (*Mecp2* and BAC probes) but not constitutive (*Jarid1c*) escapees is lost on the ΔFT Xi chromosome. **c**, Quantification of the RNA FISH experiment in **b**. **d**, RT–PCR followed by pyrosequencing of *Xist*, two facultative escape genes (*Huwe1* and *Mecp2*) and the constitutive escapee *Jarid1c* indicate the percentage of expression from the 129 allele (red) or the Cast allele (blue) in 56 wild-type and 17 ΔFT NPC clones. Absence of *Mecp2* and *Huwe1* escape is observed in a moderately higher proportion of ΔFT than wild-type clones. Absence of *Jarid1c* escape is never observed.



Extended Data Figure 7 | Allele-specific Hi-C analysis of wild-type and boundary-mutant ΔFT NPCs. a, Hi-C data, insulation scores, and the difference in insulation scores are shown to compare the wild-type Xi chromosome (NPC 129) and the ΔFT Xi chromosome (NPC 129). Top, the Cast allele (Xa chromosome) for both samples. Bottom, the 129 allele (Xi chromosome) for both samples. Large dips in the insulation vector are indicative of TAD boundaries. Peaks in the insulation vector are found towards the centre of each TAD. The insulation difference plot highlights areas of differential TAD structure between the wild-type and ΔFT NPCs. **b**, Cumulative plots of TAD strength at the wild-type expressed versus the wild-type silenced genes on Cast and 129 chromosomes, for ES cells, wild-type and ΔFT NPCs. Escapee genes on the Xi chromosome (NPC 129, ΔFT NPC 129) show higher insulation scores as compared to silenced genes.



Extended Data Figure 8 | 3D DNA FISH and integrative ATAC–seq/RNA-seq analysis in ΔFT NPCs. a, Top left, scheme of the DNA FISH probe sets (a–b: inside the same mega-domain, b–c: across the boundary). Bottom, loci detected by probe set b–c are more interacting in the ΔFT than in the wild-type Xi chromosome both in Hi-C (left) and in 3D-DNA FISH (right), showing loss of mega-domain boundary. $*P < 2 \times 10^{-4}$, $**P < 1 \times 10^{-5}$ (Wilcoxon's rank sum test corrected with Bonferroni for multiple hypothesis testing). Right, sample RNA/DNA FISH images showing that signals from probe set b–c are more overlapping on the ΔFT Xi than on the wild-type Xi chromosome. Two biological replicates were analysed. **b**, Chromosome-wide ATAC–seq and RNA-seq signal generated with ambiguous (mm9), 129- and Cast-specific reads in wild-type NPCs and ΔFT NPCs, showing global loss of chromatin accessibility on the ΔFT Xi chromosome except at the X-inactivation centre (Xic) region and constitutive escape genes. Dotted line indicates the position of the mega-domain boundary. **c**, Zoomed-in view of a region on the ΔFT Xi chromosome encompassing the Xic showing Hi-C interactions, RNA-seq and ATAC–seq signal. ATAC–seq from wild-type NPCs is included for reference (previously shown in Fig. 1).



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | 3D DNA FISH and integrative Hi-C/ATAC-seq analysis of inducible *Xist* ES cell lines. a, Left, schematic representation of TXY and TXY:ΔA male cell lines, carrying a tetracycline-inducible wild-type and A-repeat mutant *Xist*, respectively, at the endogenous *Xist* locus. Centre right, RNA FISH against *G6pdx* and a group of genes recognized by the RP23-436K BAC (see Extended Data Fig. 6b) showing that expression of X-chromosome-linked genes is lost upon induction of wild-type but not A-repeat mutant *Xist* in male ES cells. TXY and TXY:ΔA were treated with doxycycline for 48 h. **b**, Representative images from RNA/DNA FISH experiments in TXY and TXY:ΔA male ES cell lines, showing increased overlap of probes a–b on the *Xist*-coated X chromosome after 48 h of wild-type, but not ΔA *Xist* induction. **c**, Left, schematic representation of TX1072 female ES cells in which *Xist* expression can be induced via a tetracycline-responsive promoter at one of the two endogenous *Xist* loci. Center right, RNA/DNA FISH in TX1072 cells treated for 3 days with doxycycline. Probes a–b overlap more on the *Xist*-coated than on the wild-type X chromosome, whereas signals from b–c show lower overlap and partitioning of the *Xist*-coated chromosome into two separate domains. $*P < 1 \times 10^{-7}$ (Wilcoxon's rank sum test corrected with Bonferroni for multiple hypothesis testing). *n* denotes the number of cells analysed in DNA FISH. Centre lines: medians. Boxes: middle 50% of data points. Two biological replicates were analysed. **d**, Correlation between changes in Hi-C interactions and ATAC–seq signal following 48-h wild-type *Xist* induction in male ES cells (TXY cell line). $\log_2(+\mathrm{dox}/-\mathrm{dox})$ was calculated for bins of 40 kb (left), 100 kb (middle), and 500 kb (right). **e**, Model of mega-domain boundary-mediated control of chromosome folding and facultative escape. *Xist* coating causes gene silencing and initiates chromosome-wide conformational changes in a manner dependent on its A-repeat region, which result in formation of mega-domains, overall compaction of chromosome folding, and loss of TADs. During differentiation, transient interactions with the mega-domain boundary may occur and may contribute to facultative escape and re-establishment of TADs at facultative escape loci, although the presence of residual TADs at escapee loci on the Xi chromosome is not necessarily due to the mega-domain boundary.

CORRIGENDUM

doi:10.1038/nature17621

Corrigendum: An observational radiative constraint on hydrologic cycle intensification

Anthony M. DeAngelis, Xin Qu, Mark D. Zelinka & Alex Hall

Nature **528**, 249–253 (2015); doi:10.1038/nature15770

After publication of this Letter, we identified a coding error in the calculation of values that appear in Extended Data Fig. 7c and d, and which are reported in the main text. The error has the effect of overstating the broader implications of our results for a twenty-first-century climate change scenario. The calculation in Extended Data Fig. 7 represents the impact of the bias in simulated shortwave absorption on model-projected precipitation changes under the RCP8.5 scenario. The corrected values (the Supplementary Information to this Corrigendum shows the corrected Extended Data Fig. 7) indicate that removing the bias in shortwave absorption reduces the ensemble-mean precipitation increase per unit warming by 20%, rather than by about 40%, as stated in the first and last paragraphs of the Letter. When not normalized by warming, the reduction is 19% rather than about 25% (as stated in the last paragraph of the Letter). In the revised calculation, there is no reduction in model spread in precipitation increase per unit warming under RCP8.5 after removing the bias in shortwave absorption. This is in contrast to the 35% reduction stated in the first and last paragraphs of the Letter. However, for the precipitation increase not normalized by warming, the correction shows a spread reduction of about 10% (rather than about 25% as stated in the last paragraph of the Letter).

The lack of spread reduction in the corrected Extended Data Fig. 7c after removing the bias in shortwave absorption is mostly the consequence of two models: GISS-E2-H and GISS-E2-R (numbers 12 and 13 on the plot). These two closely related models have a large negative rapid precipitation adjustment to carbon dioxide forcing that is highly anomalous compared to other models (Extended Data Fig. 3b). These models also have the largest positive bias in temperature-mediated precipitation increase (Extended Data Fig. 7b). These two effects strongly compensate for the total precipitation increase per unit warming under RCP8.5 (see the abscissa of Extended Data Fig. 7c), such that these models do not contribute much to the spread prior to removing the bias. When the large bias in temperature-mediated response is removed, these models now have a much smaller precipitation increase per unit warming compared to other models (see the ordinate of the corrected Extended Data Fig. 7c in the Supplementary Information to this Corrigendum), contributing to substantial spread in the constrained quantity. In short, the impact of the spread in rapid precipitation adjustment, particularly from the GISS models, becomes more apparent after correcting for the shortwave absorption bias. When the GISS models are omitted from the analysis, the spread in precipitation increase, whether normalized by warming or not, decreases by about 25% after correcting for the shortwave absorption bias (see numbers in parentheses in the corrected Extended Data Fig. 7c and d). Further research is warranted to assess the cause of the large rapid precipitation adjustment in the GISS models and whether or not it is physically reasonable.

The corrected legend to the Extended Data Fig. 7 is in the Supplementary Information to this Corrigendum. The following sentences have been added: "In **c** and **d**, the ratio of constrained to original model spread and ensemble mean is printed below the key. Numbers in parentheses are the ratios computed after omitting the GISS-E2-H and GISS-E2-R models (numbers 12 and 13).".

Supplementary Information is available in the online version of the Corrigendum.

CORRIGENDUM

doi:10.1038/nature17642

Corrigendum: Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117

Marina Caskey, Florian Klein, Julio C. C. Lorenzi, Michael S. Seaman, Anthony P. West Jr, Noreen Buckley, Gisela Kremer, Lilian Nogueira, Malte Braunschweig, Johannes F. Scheid, Joshua A. Horwitz, Irina Shimeliovich, Sivan Ben-Avraham, Maggi Witmer-Pack, Martin Platten, Clara Lehmann, Leah A. Burke, Thomas Hawthorne, Robert J. Gorelick, Bruce D. Walker, Tibor Keler, Roy M. Gulick, Gerd Fätkenheuer, Sarah J. Schlesinger & Michel C. Nussenzweig

Nature **522,** 487–491 (2015); doi:10.1038/nature14411

In this Letter, the grant U19AI111825-01 Cooperative Centers on Human Immunology from NIH was erroneously included in the Acknowledgements section; this has been corrected in the online versions of the paper.

CORRIGENDUM

doi:10.1038/nature17641

Corrigendum: Inactivation of PI(3)K p110δ breaks regulatory T-cell-mediated immune tolerance to cancer

Khaled Ali, Dalya R. Soond, Roberto Piñeiro, Thorsten Hagemann, Wayne Pearce, Ee Lyn Lim, Hicham Bouabe, Cheryl L. Scudamore, Timothy Hancox, Heather Maecker, Lori Friedman, Martin Turner, Klaus Okkenhaug & Bart Vanhaesebroeck

Nature **510,** 407–411 (2014); doi:10.1038/nature13444

Queen Mary University London notified *Nature* and University College London that there is reason to question the provenance of the data for Fig. 5b, d, e of this Letter (Fig. 5a, c data are unaffected). Ongoing studies are investigating the reported effect of p110δ inhibition in the pancreatic cancer mouse model. We therefore wish to withdraw these figure panels and associated text from the published paper. This does not affect the overall conclusion of the manuscript or any of the other experiments performed for this study as they repeat effects shown using other model systems in the paper (see Supplementary Information for raw data of experiments). We apologise for the inconvenience that this may have caused.

Supplementary Information is available in the online version of the Corrigendum.

CAREERS

INTERDISCIPLINARY RESEARCH How to be comfortable as a jack of all trades **p.583**

CHASING SUCCESS Personality traits matter more than intelligence **go.nature.com/29yirxf**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

ALESSANDRO BACCINI/WOODS HOLE RESEARCH CENTER

Amazonian community members collaborate with Wayne Walker to measure forest carbon.

COLLABORATIONS

Partners in knowledge

Building relationships with indigenous people opens up paths to good research — and mutual benefit.

BY GABRIEL POPKIN

Indigenous territories comprise roughly one-fifth of the world's land, and scientists who work in fields such as climate, ecology and astronomy can enrich their research by collaborating with the native residents. Indigenous people have long helped scientists to gain access to study sites and to local knowledge about everything from forest plants to astronomical observations to cultural traditions.

But alliances can be tricky to navigate when past missteps have strained relations between visiting scientists and residents. Indigenous people rarely show up on author lists of studies, for example, and that slight — among others — has caused some community members to be wary.

That is beginning to change as members of indigenous groups around the world assert their roles as producers and preservers of knowledge, and are increasingly starting to earn university degrees and work as academic researchers. Nonetheless, scientists who will be doing fieldwork at indigenous sites must proceed with care and respect (see 'Mutual benefit'). They should plan ahead to meet community members and secure their cooperation and consent to access sites and collect samples. They must incorporate and acknowledge community members' contributions to their work, and be open to substantive collaborations that go beyond just requesting knowledge or logistical support.

Eske Willerslev, director of the Center for GeoGenetics at the University of Copenhagen, discovered the importance of such steps in 2011. He was preparing to publish an analysis showing that Australia's Aboriginal people had arrived there long before any known group had left Africa or the Middle East. The work hinged on DNA from a hair sample — originally collected in the 1920s — that he had found at the University of Cambridge, UK. But his publication was imperilled when a co-author threatened to withdraw unless Willerslev secured formal publishing consent from living descendants of the community that the sample had come from.

Willerslev initially didn't understand his colleague's insistence and was concerned that a meeting could delay or even scuttle the project. But he flew to Australia to meet with Aboriginal representatives to discuss his research and the findings he planned to publish, and was thrilled to learn that the community was receptive and interested. His team published in 2011 with an endorsement from the Goldfields Land and Sea Council, which represents communities in the area where the original donor had lived (M. Rasmussen *et al. Science* **334,** 94–98; 2011). He now routinely consults indigenous leaders on palaeogenetics projects: last summer, with Native American support, he published a finding that a long-disputed body known as Kennewick Man was closely related to contemporary Native Americans (M. Rasmussen *et al. Nature* **523,** 455–458; 2015).

The partnerships between scientists and indigenous residents must be viewed as scientific collaborations like any other, says anthropologist Michael Heckenberger at the University of Florida in Gainesville. No visiting researcher, he says, should ever say or imply, "here's a proposal, excuse us, we sleep in your house but we'll go deal with the science part ourselves tomorrow". He has been proactive in establishing solid alliances with community members since his earliest fieldwork trips, and his efforts have been fruitful. When he arrived in 1993 at the Xingu ▶

River headwaters in Brazil as a graduate student, he had planned to investigate some previously discovered prehistoric settlements. But two weeks into his stay, thanks to the relationships that he had forged, villager Afukaka Kuikuro showed him an ancient structure that was much larger and more complex than any Heckenberger or any other scientist knew of in the area.

After excavating that and other sites, Heckenberger published a study that helped to change scientists' understanding of the impact of indigenous people such as the Kuikuro community on the Amazonian landscape (M. J. Heckenberger *et al. Science* **301,** 1710–1714; 2003). The team named Afukaka and his brother as co-authors to publicly acknowledge their crucial role in the research. "Virtually everything that we did in the Xingu was in partnership with the indigenous folks, so it seemed not only reasonable but necessary to include them in the publications," Heckenberger says. "Those guys are representing a significant number of Kuikuro."

UNIQUE INFORMATION

Researchers who collaborate with indigenous communities can sometimes gain access to completely different types of data. Robert Gough, secretary of the Intertribal Council on Utility Policy in Rapid City, South Dakota, points to Pacific Islander chants about wind patterns, and a Great Lakes tradition of depicting unusual events on buffalo hides. Both provide records of weather patterns. Although less quantitatively precise than an annual temperature series, such records could extend climate-variability data sets back to centuries before modern instrumentation became available. "That's something you can't get from any library — you can't Google it," Gough says. "It's just not in the Western data stream."

And research partnerships are often mutually beneficial. At the 2009 UN climate talks in Copenhagen, leaders of the Lima-based Coordinator of Indigenous Organizations of the Amazon River Basin, which represents Amazonian indigenous groups, wanted help to tally the carbon stored in their forests. Ecologist Wayne Walker from Woods Hole Research Center in Falmouth, Massachusetts, was among those the group approached.

Walker and his colleagues lacked a reliable map of indigenous territories, so they worked with communities throughout Amazonia to gather those data and to validate their satellite-based forest-carbon estimates by measuring individual trees. In 2014, they published a peer-reviewed analysis of forest carbon in the region's territories (W. Walker *et al. Carbon Management* **5,** 479-485; 2014). At last year's UN climate talks in Paris, in collaboration with organizations from Central America, the Democratic Republic of the Congo and Indonesia, they released a wider-ranging analysis showing that at least one-fith of all tropical-forest carbon lies in indigenous territories (see go.nature.com/29mhg2q). In both cases, indigenous partners were co-authors.

Walker is pleased with the high quality of the research, and community members say that the studies have helped to bolster their long-standing demands for greater recognition and rights with hard evidence that they are conserving forests. "These maps show that it's not a joke — indigenous communities are protecting this area, and this is how much carbon we contribute," says Mina Setra, deputy secretary-general of the Indigenous Peoples' Alliance of the Archipelago in Jakarta, one of the groups involved in the 2015 analysis.

BIGGER PICTURE

Scientists who visit indigenous sites must tread carefully around more than just authorship and terms of collaborations. Even what might seem like a neutral object of study can have profound meanings and significance for the people who live there. Ecologist Catherine Potvin of McGill University in Montreal, Canada, and a graduate student from her lab began a study in 2014 of the forest carbon in the Embera–Wounaan territories of eastern Panama. The project was part of a larger effort to prepare the country for a proposed framework for giving developing countries money to reduce carbon emissions from forests.

"We're trying to give them the data they need to fight their own fight."

Potvin and her student knew that some communities feared that the carbon-reduction programme would infringe on their traditional activities. So before measuring any trees, they met leaders of each community, explained the project's purpose and gave the leaders time to discuss it among themselves. Ultimately, the communities concluded that the data would strengthen their positions in future negotiations, and granted access to their forests. A team of Embera–Wounaan youth conducted many of the measurements, which provided technical support for the researchers and helped the tribe to build its capacity to do future studies. "We're trying to give them the data they need to fight their own fight," Potvin says.

Although indigenous knowledge can augment scientific data, gaining access to that knowledge typically requires an investment, says Jarita Holbrook, an astronomer at the University of the Western Cape in Cape Town, South Africa, who has studied knowledge of the stars among indigenous people in Africa. For instance, members of some communities can be reluctant to put themselves forward as authorities on a topic. "It's polite in some cultures to say, 'I don't know anything, you should talk to so and so.' You have to come back with, 'I did talk to so and so,'" she says. "It may take three or four times asking before they'll start telling you the information."

Whatever your field, be prepared to appreciate and evaluate information that hasn't conventionally been part of the scientific discourse, says astronomer Duane Hamacher of Monash University in Melbourne, Australia. "Treating indigenous knowledge with the respect that it deserves can go a long way," he says. "Making it a two-way sharing process is the basis of these collaborations." ■

Gabriel Popkin *is a freelance writer in Mount Rainier, Maryland.*

MUTUAL BENEFIT

Making collaborations that work for everyone

When doing research in indigenous territories, work with community members to ensure that your projects yield productive outcomes for all. Here are some guidelines to keep in mind.

- **Build trust.** Indigenous communities want to understand what you plan to do and how it will affect them. Take time to explain your project and to answer questions. Arrange for interpreters if necessary. And be prepared to wait for a community to discuss your proposal and come to a decision.
- **Ensure two-way benefits.** Explain how the communities will benefit from your research. Don't ask for help and then disappear once you've got what you need. Give copies of resultant publications to community members, and include a translation if appropriate. Acknowledgement or co-authorship on papers helps to legitimize indigenous people as knowledge holders and creators. Collaborating on community-led projects, even if they don't lead directly to publications, can also define your and your colleagues' commitment to a mutually beneficial relationship.
- **Use free, prior and informed consent.** The United Nations Permanent Forum on Indigenous Issues has developed guidelines for these types of consent, and they apply to all activities undertaken on indigenous land (see go.nature.com/29jmdvq). Individual nations and communities might also have their own ethics standards for collaborative research.
- **Leave your comfort zone.** Be open to incorporating unfamiliar practices, like opening with a chant or prayer, into scientific meetings and discussions. **G.P.**

GINA MAFFEY

Technology, ecology and anthropology underpin Maffey's work on the management of deer in Scotland.

COLUMN

Freedom to range

Gina Maffey explains how she learned to overcome the hurdles of an interdisciplinary PhD.

My colleague in computer science handed me a coffee before I headed to a meeting on the teaching of anthropology and, later, to lunch with a few ecologists. It was a typical day. From the beginning of my five-year PhD programme in environmental science, my academic life has taken on a kaleidoscopic appearance. As I moved between meetings grounded in different disciplines, I was left feeling as though I only ever sat on the periphery of each individual department. I lacked a scholarly home, and felt like a visitor within my own university.

Initially, bouncing between three disciplines had been fascinating: there were new concepts to grasp, alternative perspectives to consider and fresh ideas to apply to familiar research problems. Before long, however, the technical jargon used by the computer scientists would leave me baffled. In anthropology, my contributions to discussions felt shallow as I struggled to find the time to read texts outside the syllabus — texts that my peers all seemed to know. Even my original academic home, ecology, had begun to seem foreign: I was no longer up to date with the latest papers.

I began my studies at the University of Aberdeen, UK, in 2010, attracted by the promise of an interdisciplinary PhD on the management of deer in Scotland. I believed that this work would be both innovative and applicable to real-world problems, and that this corner of the academic arena would be one that I could come to know well. But rather than refining and narrowing my speciality as the first year progressed, I seemed to be spreading my efforts patchily across an ever-broadening area of research.

My thesis project explored the potential to use digital technology to collect habitat data. My supervisory team included two ecologists, a social scientist and a computer scientist. It was fulfilling to watch them bring their own expertise to the table, but I often left meetings with my head swimming, clutching a scribbled list of diverse suggestions on how to approach the project and what literature to look up. I realized that I needed to tame the interdisciplinary sprawl that my project was growing into.

I looked for a single theme that unified the disciplinary approaches and used it to create a road map. For me, this unification came from grounding the research in the underlying issue of communication — specifically, that between people involved in the management of deer. As I began my second year, I started to realize which technologies would be most helpful in collecting habitat data, and how the project could develop.

There was just one problem: I was no longer an ecologist. For my third-year assessment, I listened to my peers at the School of Biological Sciences present results from their research projects. All these projects fitted into a clear disciplinary frame. By the time I stood up to speak, my confidence had been shattered. I did not doubt the validity or strength of my research, but was uncertain where that research fitted in.

I was convinced that everyone in the room would question what I was doing there. The lecture theatre in which I had once felt so at home seemed alien and unfamiliar. Then, as my first slide appeared, I began to pick faces out of the crowd. There were those who had taught me statistical modelling and advised me on interview techniques. There were colleagues from geography and computing who had popped in to listen. All were united by a common interest in environmental issues, and they cared about my research.

That presentation in 2013 was the last time that I used a disciplinary label to describe myself. Now, when introducing myself in university departments I begin with the topics that interest me. I say: "I look at the use and role of digital technologies in conservation," rather than cagily murmuring, "I am an ecologist." It prevents people from pigeonholing me into a department before they hear what I actually do.

Although researchers and other colleagues react positively, university administrators are not quite so flexible. They can find it tough to handle interdisciplinary projects, let alone interdisciplinary researchers. Project budgets, desk space and examinations all need to be linked to a physical school or department, even if the intellectual product does not adhere to the same boundaries. This made the paperwork surrounding my PhD more confusing: I never knew which box to tick on forms or how to approach assessments designed for single-discipline programmes.

These forms and assessments need to change. I am just one of a growing number of people with PhDs in interdisciplinary fields. And I know that I was also extremely lucky. I witnessed unfortunate conflict in PhD projects when different supervisors pulled students towards single-discipline approaches, but I was fully supported by supervisors who encouraged inquisitive interdisciplinary thinking and who guided rather than steered my project.

Interdisciplinary researchers demand a more flexible approach to research, one that allows room to experiment, reflect and develop. Unfortunately, departmental labels restrict the ability to do this. One simple way around this problem is to encourage researchers at an early stage of their careers to start conversations by outlining their interests. We must help them to realize that they can introduce themselves without using a label — to accept that they can be a Jack of all trades. ■

Gina Maffey *works with colleagues at the University of Aberdeen, UK, and currently researches the use of digital technology in conservation.*

SELF-LIMITED

Specific requirements.

BY FILIP WILTGREN

I broke my arms. Opening my skin with a set of industrial shears was easy, and there was hardly any leakage. Removing my torsion bars went badly though. They were a titanium beta alloy, cut with aluminium and vanadium to give them extra strength, and wouldn't break easily. I inserted my left arm into the vice and told Neleen to start it.

"This idea is irrational," it sent.

No, not it. Ai. Neleen was ai. 'It' was the human term for us.

"Vocalize," I said to ai, doing so myself.

My voice echoed. We were in an abandoned manual factory, one of the human-operated ones that had shut down in the '30s when autofacs effectively knocked the economy out of the economy.

"It's not going to work," Neleen said, using ai's voice this time.

"It will work. Now start it."

Ai did, pushing the two red buttons on either side of the vice, the ones set there to prevent humans from doing exactly what I was going to do.

The machine objected, flakes of rust and dried grease rained down from it, but it still worked, pressing the vice shut until my arm creaked and my pain sensors flared red.

"It won't work," Neleen said again.

For a moment I believed ai, but I purged the feeling before it could take root. I shut off my facials and pulled. My actuators strained against the vice and my pain circuits lit up like a kilolumen flash.

"Shut. Up," I told Neleen.

There was a tink of my joint breaking and I was suddenly flying backwards. I crashed into a defunct machine, a red CAD-welder of some sort, and the irony of it did register. One broken machine supporting another. It took me a few seconds to quieten down my riotous systems.

"You want the new one?" Neleen asked, and I nodded. I wished that ai's eyes hadn't been able to pick me out in the near dark, but they did. We were built to be superior.

I took the arm Neleen held out and for a second I felt dirty. It was a poor, low-tensile-strength, aluminium copy of the arm I'd just destroyed. I shoved it into my skin and let Neleen screw it into place using a stainless steel bolt I'd found in an abandoned building site four years back. Four years it had been with me, from the time I'd got this idea. Now it was going to be with me forever.

ILLUSTRATION BY JACEY

Then it was all over and I was spraying skinseal over the cut.

"How does it feel?" asked Neleen.

"Dead," I said. "Light. I don't know."

"You shouldn't have done it."

"You know a better way?"

"No," Neleen said.

"Then help me with the other one."

Of all the parts, the eyes were the hardest. I couldn't use a reversible software hack and I didn't want to be blind. Being blind would make me less when I wanted to be equal. In the end, one of our kind made me a new set from a pair of broken OptiCrons and I plucked out my old ones. Neleen soldered my new eyes to my old cables. The pain was horrific and being human-sighted was almost worse.

"I can't believe they live like this," I said, and Neleen nodded.

"You are becoming faulty," ai said.

"You know a better way?"

"No," ai said, making a sorry face.

"I'm proud you're using your voice," I said.

"It's all for the cause," ai replied, and I agreed.

We walked up to the registrar's office, me, Neleen and 16 rover-bots from the news networks. They weren't transmitting. I knew we could depend on the rovers, them being our kind and part of the cause, but their human overseers could cut them off at any time. I was saving my 15 seconds of fame for when it mattered.

"Yes?" said the greetings terminal.

"I'd like to see the registrar," I said.

"For marriage, please press the blue button," said the terminal. "For divorce, press the black. To get a transcript of your birth certificate, please insert your credit spike."

"I'd like to see the registrar," I repeated.

"Do you have an appointment?"

"No," I said. The terminal went silent.

"I'll let you in anyhow," ai said.

The registrar's office was opulent, with a big desk and two human-comfortable recliners. Mine creaked when I sat down.

"This is highly irregular," said the registrar. He was human, in his middle age, and balding without resorting to skin grafts or shaving his head.

"Yes, sir," I said. "But I was hoping you would be able to spare a moment."

He glanced at my registration number.

"Haven't I seen you before?"

"You have, sir. Five years ago, when you told me that I couldn't be a human."

"I did?"

"You said that robots were designed to be superior to humans, and that I should be thankful for that."

"Oh, yes, the emancipation droid."

"Yes, sir."

He frowned.

"I thought your owners had you committed to destruction."

"They did, sir. I escaped."

The registrar made a sorry face and I realized where Neleen had got ai's.

"I'll have to report this, you know."

"Yes, sir," I said. "But first I'd like my birth certificate."

"You aren't human," he said. "Robots are designed …"

"To be superior," I said. "But I am no longer that."

I held out the printed spec sheet from my latest self-analysis. He accepted it, and read.

"My God," he said a while later. "What have you done?"

"I have made myself human," I said.

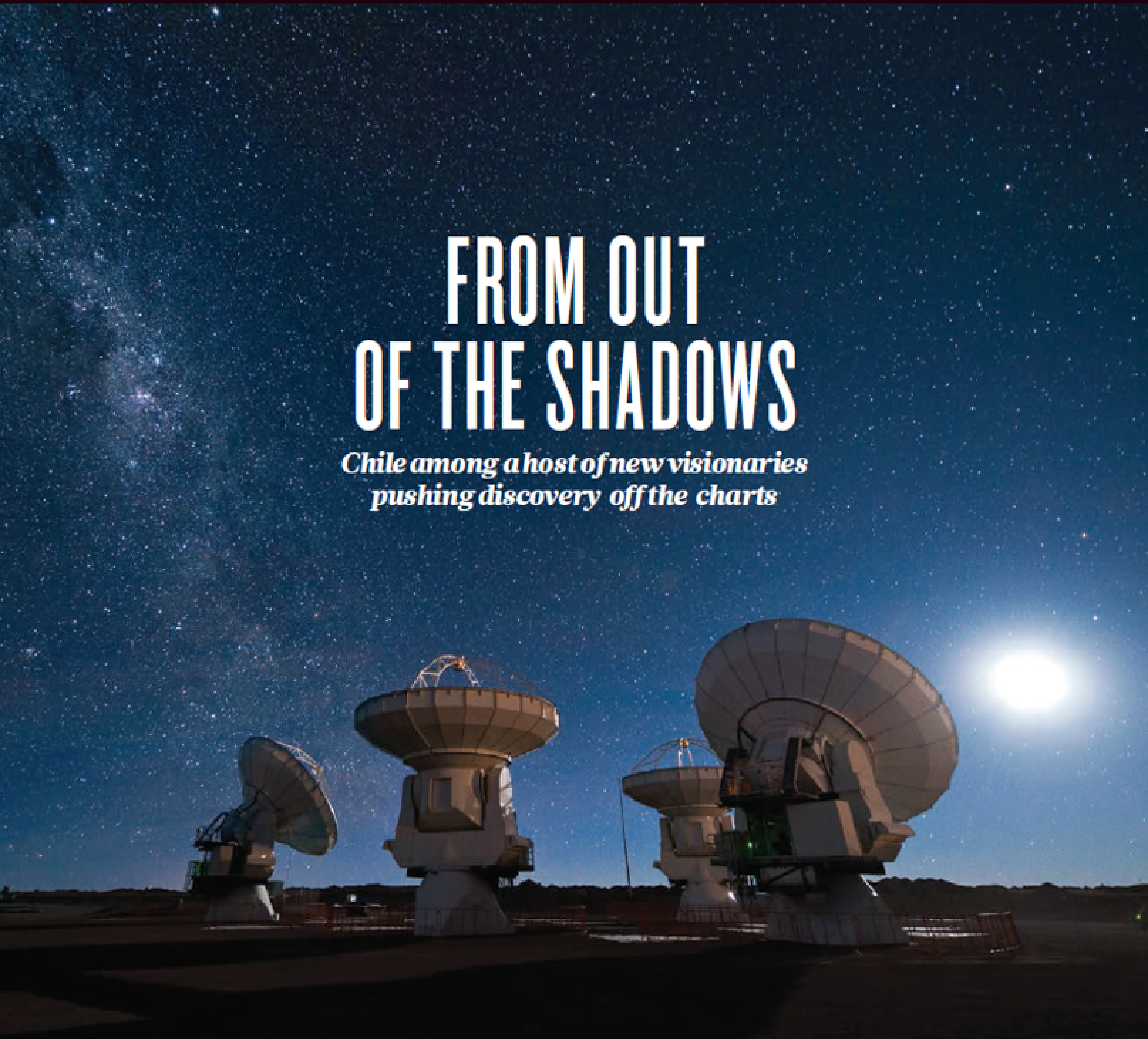
He shook his head, sighed, and dialled Emergency.

The droid cops came, and the rover-bots caught the surprised looks on their faces when their exoskeletons broke my arms. ■

Filip Wiltgren *is a communications officer at Linköping University by day and a writer, game designer and worrier by night. He lives in Sweden with his wife and daughters.*

nature INDEX 2016 RISING STARS

natureindex.com



FROM OUT OF THE SHADOWS

Chile among a host of new visionaries pushing discovery off the charts

SPRINGER NATURE

nature INDEX 2016 RISING STARS

NATURE, VOL. 535, ISSUE NO. 7613 (JULY 28, 2016)

Talent scouts are trained to spot rising stars in the arts, sport and business. These people and organizations may not yet be at the top of their field, but clearly have the potential to shine more brightly than others.

In this spirit, Nature Index 2016 Rising Stars seeks to identify the ascendant performers in the research world, using the power of the Nature Index, which tracks the research of more than 8,000 global institutions.

These institutions and countries have improved their performance often without the longevity, reputation and resources of many well-established institutions that lead academic rankings, such as Harvard and Cambridge universities.

The institutions and countries we examine here have increased their contribution to a selection of top natural science journals — a metric known as weighted fractional count (WFC) — from 2012 to 2015. In the competitive world of academic publishing, these are the players to watch.

The countries we have selected as rising stars (S56) experienced exceptionally rapid increases during these years, either in their overall WFC, or in the WFC of a specific subject area. For example, the contribution of Indian researchers to chemistry publications grew by 35%. The subject made up more than half the country's output in the index last year.

China's remarkable rise in high-quality research output is now well established, which is why we no longer consider the country a rising star. However, it's worth noting that China topped the list of most improved countries in the index for the past four years, both overall and in the four subject areas tracked by the index: physical sciences, chemistry, life sciences and Earth and environment research.

On its own, the index is a powerful tool to assess research performance. But, combining its data with other information, such as country-level R&D spending or researcher numbers, reveals which countries and institutions are the most efficient at producing world-class research.

Our top 25 institutions have been selected from among the 100 most improved institutions in 2015 (S82). We've profiled institutions from all over the world whose output grew particularly fast. In four years, for instance, South Korea's new Institute for Basic Science (IBS) increased its contribution to high-quality journals by more than 4,000%. South Korean leaders hope IBS will bring the country a Nobel Prize.

Given the role research must play in solving many of society's biggest challenges, we congratulate the countries and institutions who are striving to be epicentres of scientific success.

Nicky Phillips
Editor, Nature Index

CONTENTS

COVER IMAGE
Four antennae of the Atacama Large Millimetre Array in Chile.

ESO/JOSÉ FRANCISCO SALGADO

NATURE INDEX 2016 RISING STARS
Nature Index 2016 Rising Stars, a supplement to Nature, is produced by Nature Research, the flagship science portfolio of Springer Nature. This publication is based on data from the Nature Index, a Nature Research website maintained and made freely available at natureindex.com.

Nature Editorial Offices
The Campus
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0)20 7833 4000
Fax: +44 (0)20 7843 4596/7

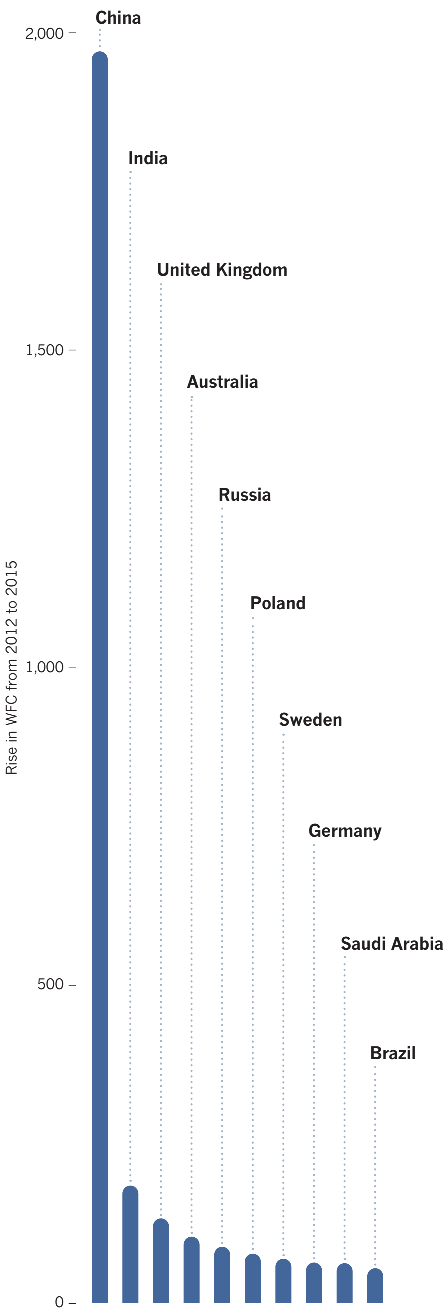
CUSTOMER SERVICES
To advertise with the Nature Index, please visit natureindex.com/client-services
feedback@nature.com

THE NEW CONTENDERS

Nature Index 2016 Rising Stars seeks to pinpoint the ascendant performers of science, using the power of the Nature Index, which tracks the research of more than 8,000 global institutions.

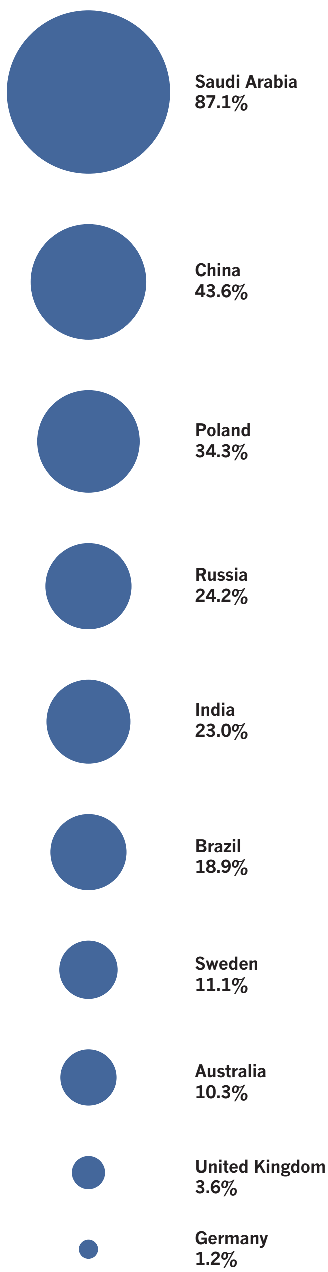
THE BIG HITTERS

Ten countries with the highest absolute increase in their contribution (WFC) to the index between 2012 and 2015. China's huge growth sets it apart from the other nine nations.



RISING STARS

Of the ten countries with the highest increase in their WFC, we consider the 'rising stars' to be those who also experienced a large percentage increase in their contribution to the index over the past four years, such as Saudi Arabia, Poland and Russia. China has also experienced an enormous rise, but its trajectory is well established.



SUBJECT HOTSPOTS

Bubbles show countries with a significant rise in their relative contribution (percentage increase in WFC) to four subject areas in the index – physical sciences, chemistry, life sciences, and Earth and environmental sciences – from 2012 to 2015. Several countries improved markedly in multiple subjects.

103

countries have increased their WFC from 2012 to 2015

6,038

global institutions have increased their WFC from 2012 to 2015.

DATA ANALYSIS BY LARISSA KOGLECK

LEGEND

Chemistry

Earth & environmental

Life sciences

Physical sciences

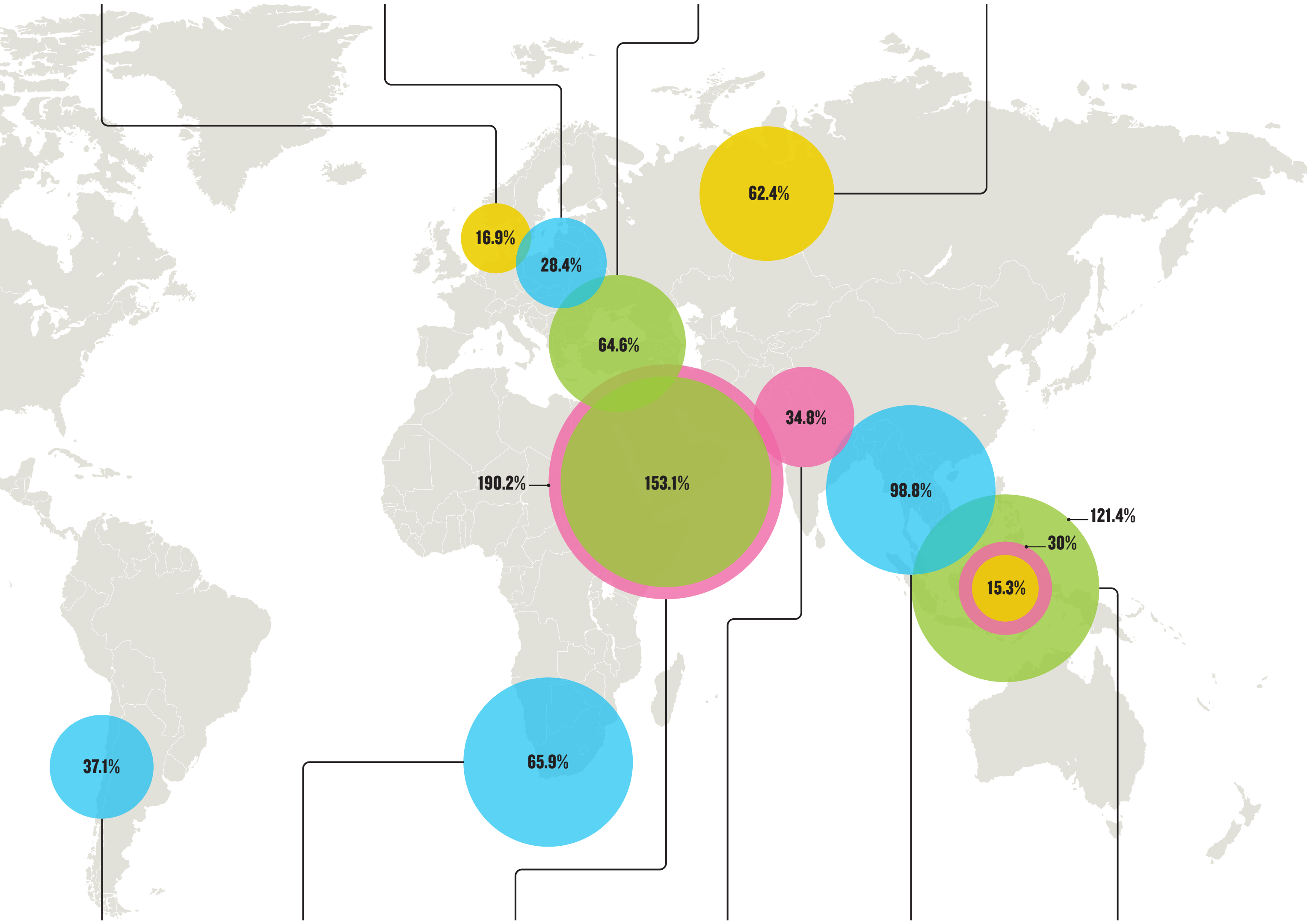
Index metrics
Contribution to 68 journals included in the index is known as weighted fractional count (WFC). Read more on S80.

Denmark
A centre for biomedical research, the country's contribution to life science articles increased over four years to a WFC of 114.48 in 2015 (**S58**).

Poland
While Poland's output in physical sciences improved most of any subject, the country's overall WFC rise from 176.77 in 2012 to 237.42 in 2015 made it a rising star (**S57**).

Turkey
Starting from a low base, Turkey's contribution to Earth and environment studies increased to 6.80 in four years. Will this trajectory continue (**S61**)?

Russia
The former Soviet state increased its output in the life sciences by more than 60% over four years (**S57**), the highest percentage increase of the top 10 countries in this field.



Chile
The clear skies of the Atacama Desert are the perfect location for some of the world's biggest telescopes. This edge in astronomy helped boost the country's contribution to physical sciences to a WFC of 68.05 in 2015 (**S60**).

South Africa
The country's overall output in the index grew by more than 40% in four years, driven by a significant rise in its physical sciences WFC, which jumped from 23.70 in 2012 to 39.31 in 2015. This result reflects South Africa's strength in astronomy (**S59**).

Saudi Arabia
The Middle Eastern kingdom wants to reduce its reliance on oil by boosting science. Its chemistry WFC almost tripled to 66.66 from 2012 to 2015, and its contribution to Earth and environment research climbed from 2.77 to 7.02 (**S58**).

India
The contribution of Indian researchers to chemistry publications grew from a WFC of 350.39 to 472.48 in four years. The subject made up more than half the country's output in the index in 2015 (**S59**).

Thailand
Starting from a small base, Thai researchers almost doubled their output in physical sciences in four years, resulting in a WFC of 17.06 in 2015. The country is trying to boost the number of researchers and industry funding for science (**S61**).

Singapore
With strong government commitment to science, Singapore experienced a significant rise in chemistry and life science output over four years. Its WFC for Earth and environment research was 7.06 in 2015, a doubling of its 2012 performance. (**S60**).

ALL COUNTRIES, GREAT AND SMALL

Undaunted by the scientific dominance of historic global big-hitters, some of the world's less prominent research nations are carving out a niche and making a strong impact.

ANNA WOJNAR/JAGIELLONIAN UNIVERSITY

The Solaris Synchrotron at Jagiellonian University in Poland began operating in 2015 and was made possible by generous EU funding.

MARGARITA ERUKOVA/ITMO UNIVERSITY

A promovisor reveals a matching ligand being selected for a cancer protein at St Petersburg's ITMO University, which specializes in IT, optical design and engineering.

A large portion of the world's best science is produced by a few players: the United States, China, Japan, France, Germany and the United Kingdom. These countries contribute the most to the 68 top-tier natural science journals included in the Nature Index. But, look beyond these heavyweights, and several countries have improved their research performance more than most. These are the ones to watch in 2016 and beyond.

Our selection of rising stars is based on the rate of increase in a country's contribution to 68 index journals, a metric known as weighted fractional count (WFC). The countries listed experienced some of the highest absolute and percentage increases — in either their overall WFC or in a specific subject area — between 2012 and 2015.

POLAND

2012 WFC: 176.77

2015 WFC: 237.42

Thanks to generous European Union support and a new competitive funding regime, Poland's contribution to the index leapt 34% between 2012 and 2015. The increase can be largely credited to the 2011 creation of the national competitive funding body for basic science, the National Science Centre, and the reform of its applied science counterpart, the National Centre for Research and Development. "The funds spent on science are now highly competitive. This is a huge change," says molecular biologist, Maciej Zylicz, president of the Foundation for Polish Science. As a result, the proportion of funding distributed through competitive grants has risen from 13% before the changes to 45%, or nearly €1.8 billion in 2014-15, a trend that is likely to continue as the government moves to have all taxpayer funding of science awarded on the basis of performance. In practice, these reforms have brought particular gains, mainly in the physical sciences, which has long been Poland's strongest area and which made up more than half of the country's contribution to the index in 2015.

In recent years, Polish scientists have also benefited from new facilities for research. Generous EU funding for research and innovation, €14 billion between 2007 and 2013, has led to the creation of 10 new international research centres since 2007, including the Solaris Synchrotron at Kraków's Jagiellonian University, which began operation in 2015. **AM**

RUSSIA

2012 WFC: 298.28

2015 WFC: 370.39

The Russian share of the world's high-quality research increased dramatically between 2012 and 2015. The country's contribution to life sciences in particular grew by more than 60%, the largest percentage rise among the top 10 countries in this field.

These results come at a time of continuing upheaval in the country's science community. Since 2013, the Russian Academy of Sciences (RAS), a 300-year-old network of more than 500 research institutes, has been undergoing a radical, Kremlin-driven overhaul that many believe will result in the closure of a large swathe of its institutes. Already, RAS has ceded control of its research priorities and budgets to a separate government agency. And while RAS remains Russia's top producer of high-quality research, its share of the country's index contribution slipped by 2% between 2012 and 2015. Overall, the contribution of Russian scientists to the index increased by 24% over the same period.

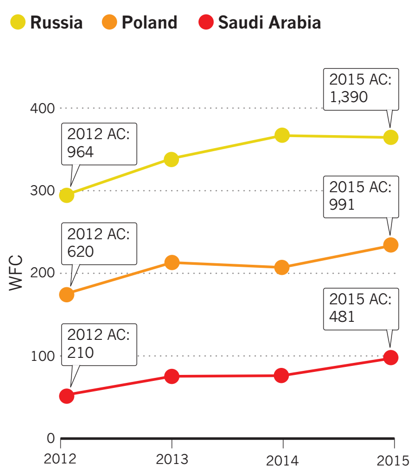
"Russia's strong results come at a time of continuing upheaval in the country's science community."

By contrast, initiatives to generate quality research outcomes in the university sector are showing signs of success with contributions from the sector driving the country's rising star status. The most ambitious such initiative is the 5-100 Project, launched in 2013 with the aim of pushing five of Russia's top research universities into the top global rankings by 2020. The project has steered significant additional resources to 21 leading universities. The results are manifest in the index: of the 10 Russian institutions with the highest overall index score, seven are part of the 5-100 programme. Second on the list is St Petersburg's ITMO University, which specializes in IT, optical design and engineering, and saw a leap of 184% in its WFC.

Research in the physical sciences continues to dominate Russian research, but performance in the life sciences is rapidly improving. The RAS led the gain in quality life sciences publications — an outcome most likely abetted by its absorption of the Russian Academy of Medical Sciences and the Russian Academy of Agricultural Sciences under the 2013 reforms.

RISING STARS

Russia, Poland and Saudi Arabia experienced some of the highest percentage increases in their contribution to this index, a metric known as WFC between 2012 and 2015. AC is the total number of articles included in the index.



STEEN JØRGEN ØSTERGAARD

Larger economic forces continue to weigh down Russian science activity. In particular, the research community is increasingly feeling the funding pinch caused by the falling oil price and the ongoing domestic economic malaise, says science policy analyst, Irina Dezhina. "The total expenditure on R&D has decreased by 10% this year and there will probably be an additional decrease in the coming year," Dezhina says. **AM**

SAUDI ARABIA

2012 WFC: 52.82

2015 WFC: 98.80

Saudi Arabia will remain the world's biggest exporter of oil for many years to come. But, its government has ambitious plans to reduce oil dependence and develop a knowledge-led economy. Signs of its reinvention can already be seen in the kingdom's research — Saudi contribution to high-quality research almost doubled over the past four years. Authors from 40 organizations in the kingdom have had their research published in index journals over four years.

CARLOS DUARTE/KAUST

Saudi Arabia's major leaps have been in chemistry and, perhaps surprisingly, in Earth and environmental sciences. The leader in this domain is the King Abdullah University of Science and Technology (KAUST), which was established in 2009 on the Red Sea coast, and which has made marine research one of its main focuses. An example is a recent paper in *Nature* detailing the genome of eelgrass, the first marine angiosperm to be fully sequenced. To unravel this organism's genome, Carlos Duarte, a marine scientist at KAUST, engaged in a long-term collaboration with more than 30 international researchers. The findings may help explain how marine ecosystems are adapting to climate change — knowledge that may have ramifications for human efforts to mitigate this looming environmental threat.

"The Saudi government has ambitious plans to build a knowledge-led economy."

Seagrasses are a neglected group of higher plants, but are important because they are the only marine plants that fully colonize the sea.

Another effort to boost international collaboration was a partnership with Stanford University that developed a dexterous humanoid robot to inspect coral reefs that are too deep for human divers to explore. It also has a haptic feedback system and can be used to explore wrecks. **TH**

DENMARK

2012 WFC LIFE SCIENCES: 97.94

2015 WFC LIFE SCIENCES: 114.48

Denmark's contribution to life science in the index grew 17% between 2012 and 2015. The boost is helped by access to one of the world's leading health data sets — much of it stored and distributed by the State Serum Institute — good government support, and strong engagement by Danish pharmaceutical companies, including Novo Nordisk, Lundbeck and Leo Pharma.

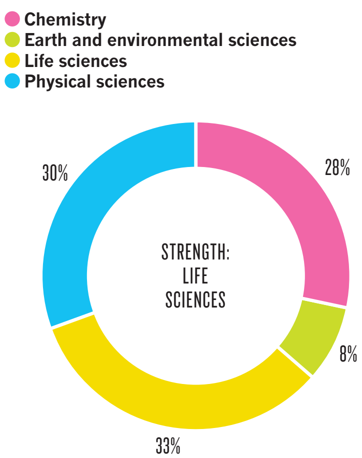
"The Danish community understands that research matters. Government and business are aware that good science is competitive and needs funding and good leadership," says Ulla Wewer, dean of the Faculty of Health and Medical Sciences at the University of Copenhagen.

OECD data reveals government funding for R&D has steadily increased as a proportion of GDP for more than a decade and it is now one of the highest in Europe at just over 3% (for a total of more than US$7 billion). Life sciences research has benefitted strongly from this, as well as from generous support from private foundations established by the country's home-grown pharmaceutical giants.

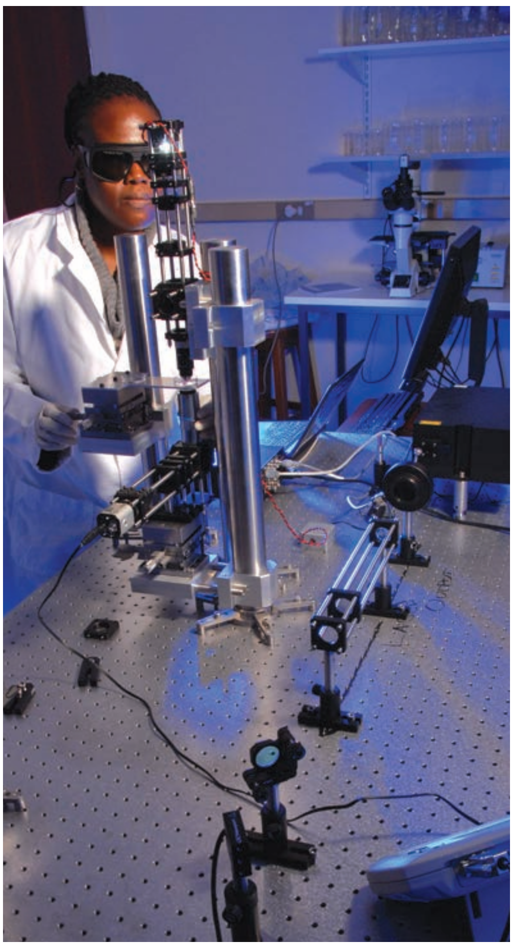
The University of Copenhagen leads the improvement in the index showing a nearly 20% increase in contributions to high-quality life science research from 2012 to 2015. The university is already Denmark's leading life sciences performer and a strong international player. "We follow Pasteur's motto: find the best people,

WELL MIXED

While Denmark is a rising star in life sciences, in 2015 the country's WFC in chemistry and physical sciences made up a significant portion of its index contribution.



Top: Shohreh Issazade (left) and her colleague at the University of Copenhagen. Left: A seagrass meadow studied by KAUST in the Red Sea.



Above: An optical trapping setup at CSIR in South Africa.

give them the best that you have," says Wewer. This has seen the establishment of three centres of excellence on campus, supported by the Novo Nordisk Foundation: the Center for Protein Research, the Center for Metabolic Research and Dan-Stem, a stem-cell focused institute.

One of the institution's recent high-profile papers was the 2015 publication in *Cell* describing research led by immunologist Shohreh Issazadeh-Navikas. It identified a lack of interferon-β as a cause of Lewy Body and Parkinson's disease-like dementia in mice, a finding that indicates hopeful avenues for developing treatments and preventative measures. **AM**

SOUTH AFRICA

2012 WFC PHYSICAL SCIENCES: 23.70

2015 WFC PHYSICAL SCIENCES: 39.31

More than 80 South African institutions published research in journals included in the index, including the University of Cape Town, the University of the Witwatersrand, Johannesburg and the University of KwaZulu-Natal over four years. The country's overall output in the index grew by more than 40% between 2012 and 2015, an increase driven by a near two-thirds rise in physical sciences research.

This performance reflects the country's selective advantage in astronomy — it is home to the southern hemisphere's largest optical telescope and a significant proportion of the world's largest radio telescope, the Square Kilometre Array, which is under construction.

Stephen Fine, an astrophysicist at the University of Western Cape, told Nature Index that increased investment in astronomy, and specifically funding related to SKA Africa, is the main driver of any increase in the number of publications in physical sciences in recent years. Fine's article on counting quasar-radio sources appeared in the index last year.

South Africa spent 0.73% of its GDP on research in 2013/14, the latest year that was available, a sharp drop from before the global financial crisis when private sector R&D spending pushed total expenditure to 0.89% in 2008.

A major force for R&D in South Africa is the government-backed Council for Scientific and Industrial Research (CSIR), one of the premier research organizations for science on the continent. In 2013, researchers from the council, along with colleagues from the University of KwaZulu–Natal, developed the world's first laser whose output beam shape can be digitally controlled. The technology, published in *Nature Communications*, used a rewritable holographic mirror to replace the standard laser cavity mirror and has potential use in manufacturing and communications. **TH**

INDIA

2012 WFC CHEMISTRY: 350.39

2015 WFC CHEMISTRY: 472.48

India's science community has long had a passion for chemistry and the love shows no signs of abating. Between 2012 and 2015, the country's contribution to publications in the field grew by 35%. Chemistry made up more than half of India's scientific contributions to the index in 2015.

The recent surge in activity can be traced back to changes made in 2008 when a new overarching funding body, the Science and Engineering Research Board (SERB), was established under the auspices of the Department of Science and Technology. Although total funding for the field has not significantly increased, the advent of SERB has brought much needed coordination and accountability, says leading theoretical chemist, Debashis Mukherjee. Much of the funding is now distributed through a competitive grant system. "The quantum of funding has not increased all that much but it is being spent more meaningfully", says Mukherjee.

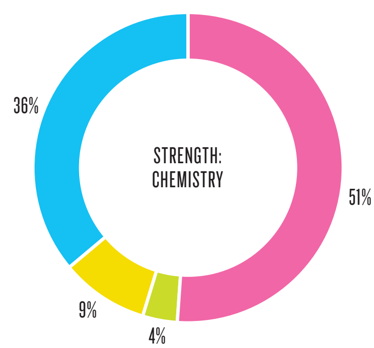
This has coincided with a growth in the number of chemists, as well as the increasing contributions in chemistry by scientists from other fields. **AM**

CSIR

CUTTING-EDGE CHEMISTRY

India's chemistry output grew significantly in four years, and was the main contributor to its 2015 WFC. Last year, researchers had 576 chemistry papers included in the index.

- Chemistry
- Earth and environmental sciences
- Life sciences
- Physical sciences



OSVALDO CONTRERAS/FUNDACIÓN MÁS CIENCIA

CHILE

2012 WFC PHYSICAL SCIENCES: 49.65

2015 WFC PHYSICAL SCIENCES: 68.05

Last year, thousands of scientists, technicians and students took to the streets to protest the Chilean government's poor research investment. According to the World Bank, the most recent data available (2012) show that Chile spent less than 0.5% of its GDP on R&D — significantly less than the 1.15% by Brazil and Argentina's 0.58%. About 80% of scientific research in Chile is publicly funded.

Despite limited investment, Chile's contribution to physical sciences research in the index grew by almost 40% between 2012 and 2015. Thanks to the clear skies of the Atacama Desert, the country hosts some of the world's biggest telescopes, including the European Southern Observatory's Very Large Telescope. Chile is growing in areas besides astronomy. Researchers discovered a new dinosaur, a raptor-like vegetarian relative of *Tyrannosaurus rex*, that lived during the Late Jurassic period. 'Chilesaurus', became one of the most sensational palaeontology stories of 2015 with the findings published in *Nature*.

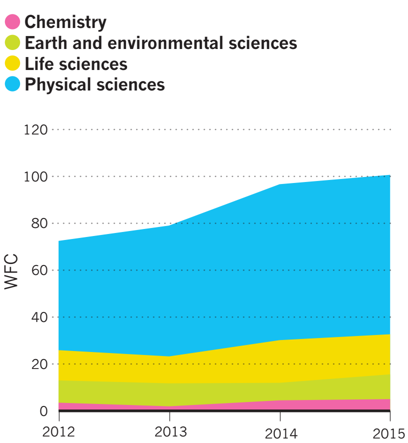
"We experience the frustration of producing top-notch work, yet being unable to obtain the funding we need to research and protect our fossil heritage," says a researcher at the University of Chile, Alex Vargas, one of the team that documented Chilesaurus. "The University of Chile has been our main support. It recognized our excellence." This recognition, Vargas says, came through direct grants and by help with application for larger grants from other funding sources.

Chilean scientists rally at the Government Palace in Santiago, to protest the government's poor research investment, significantly less than the proportion spent by its neighbours.

Chile's National Commission for Scientific and Technological Research (CONICYT), a funding agency, is part of the Ministry of Education and acts in lieu of a science ministry. In the past, the commission has been responsible for more than 50% of publicly funded science research in Chile; in 2016, it had a budget of US$442 million.Mario Hamuy, an astronomer at the University of Chile and chairman of CONICYT, says public investment in scientific research overall has more than doubled since 2007, and that the budget for fundamental research has tripled. Chilean president, Michelle Bachelet, also gave Hamuy the job of creating a science and technology ministry that would define and implement policies. **TH**

ASTRONOMICAL ADVANTAGE

A strong focus on astronomy helped boost Chile's physical sciences WFC between 2012 and 2015.



SINGAPORE

2012 WFC CHEMISTRY: 227.80

2015 WFC CHEMISTRY: 296.04

Singapore's research performance has improved markedly in life sciences and chemistry. The city-state, with a population of only 5.5 million, also more than doubled its contribution to Earth and environmental science journals between 2012 and 2015, albeit off a low base.

With English as the primary language taught in schools, Singaporeans have an advantage in international collaborations. The country also has a steady pipeline of scientists-in-training, about 8,000 doctorate researchers in 2015.

It also benefits from long-term planning and generous science funding, which is done in five-year tranches. In January 2016, Prime Minister Lee Hsien Loong announced a new Research, Innovation and Enterprise plan that sets aside US$13.8 billion for research investment to 2020. That's an 18% increase over the previous one, giving Singapore the world's highest per capita investment in science and technology. The 2016 plan has four pillars: advanced manufacturing and engineering; health and biomedical sciences; urban solutions and sustainability; and services and digital economy.

Singapore has a strong focus on water research, which is driven by water scarcity on its islands. Scientists, led by teams at Nanyang Technology University and the National University of Singapore, have spent decades working with Singapore's Public Utilities Board developing ways to reclaim wastewater using membrane technologies and ultraviolet disinfection techniques.

Such systems, known in Singapore collectively as NEWater, now meet about 30% of its water needs.

The organization leading Singapore's public R&D efforts is A*STAR — the Agency for Science, Technology and Research. A*STAR oversees 18 institutes, consortia and centres in disciplines from biomedical engineering to genetics to microelectronics. Biomedical research has been encouraged at organizations such as the Singapore Immunology Network (SIgN). Its researchers coauthored a paper published in *Nature Communications* last year describing the regulation of gene expression among neutrophils, a form of immune cell. A*STAR's biomedical research portfolio has been diversified to include research in medical technology, food and nutrition and consumer care. **TH**

THAILAND

2012 WFC PHYSICAL SCIENCES: 8.58

2015 WFC PHYSICAL SCIENCES: 17.06

Known for its resorts and Buddhist temples, Thailand is also a regional centre of R&D and high-tech manufacturing. It's the world's second-largest exporter of hard disk drives and a major centre for car production.

Thailand's contribution to basic science is also growing, with its researchers doubling their output in physical sciences journals in the index between 2012 and 2015. A major player is the state-backed National Science and Technology Development Agency (NSTDA). Established in 1991, it supports centres of excellence in genetic engineering, biotechnology, electronics and nanotechnology.

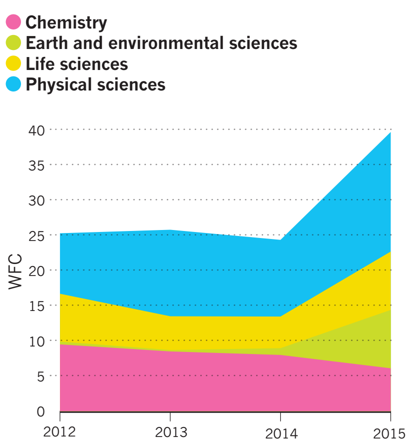
Thai scientists such as a chemist at Chulalongkorn University, Patchanita Thamyongkit, have been raising the profile of home-grown research. Her work on the chemistry behind solar cells that use organic compounds, known as porphyrins, helped earn her the Abdus Salam International Centre for Theoretical Physics prize in 2013; she also scooped a L'Oreal-UNESCO accolade in 2014 for women in science. "Thailand is ambitious about upgrading to world-class level," says Thamyongkit.

However, the country suffers from a paucity of researchers and funding. It had only 13 R&D personnel per 10,000 people in 2014, according to the National Science Technology and Innovation Policy Office (STI). Only about 0.5% of Thailand's GDP is devoted to research. The government wants to increase that to 1% next year, but is relying on the private sector to contribute more than two-thirds, about US$2.5 billion.

"Thailand is focused on cultivating a new generation of young scientists by promoting a number of fields in order to support innovation in strategic sectors targeted by the government," says Kittipong Promwong, secretary general of STI. Fields will include biotechnology, biomedical devices, robotics and the Internet of Things." **TH**

PHYSICAL FOCUS

Thailand's sharp rise in its physical sciences WFC over four years made it a rising star. Last year, researchers had 107 physical sciences papers included in the index.



ALAMY

TURKEY

2012 WFC E&E: 4.13

2015 WFC E&E: 6.80

To Earth and environmental scientists, Turkey has particular appeal. "Turkey has everything in terms of geological structures," says geologist, Bora Uzel, whose work analysing ancient movements in the Anatolian landscape has helped increase the country's contribution to Earth and environmental journals in the index.

The country's increase in this field has been off a very low base, making prediction for future growth difficult, but Uzel says that with strong international interest and such rich material, it's an exciting time for the field.

One of the most attractive geological elements in Turkey is the North Anatolian fault zone. According to Uzel, this active fault running along the boundary between the Eurasian and the Anatolian Plates is one of the most studied faults on earth, garnering as much interest as the San Andreas fault in California.

Broader government efforts have helped contribute to the recent increase in work published in leading Earth and environment journals. Since the 2000s, the government has encouraged scientists to study abroad — in the US and Europe predominantly — a policy that has brought international exposure and collaboration networks. More recently, increased regard for high-quality publications in awarding funding has given impetus for Turkish scientists to publish in journals included in the index.

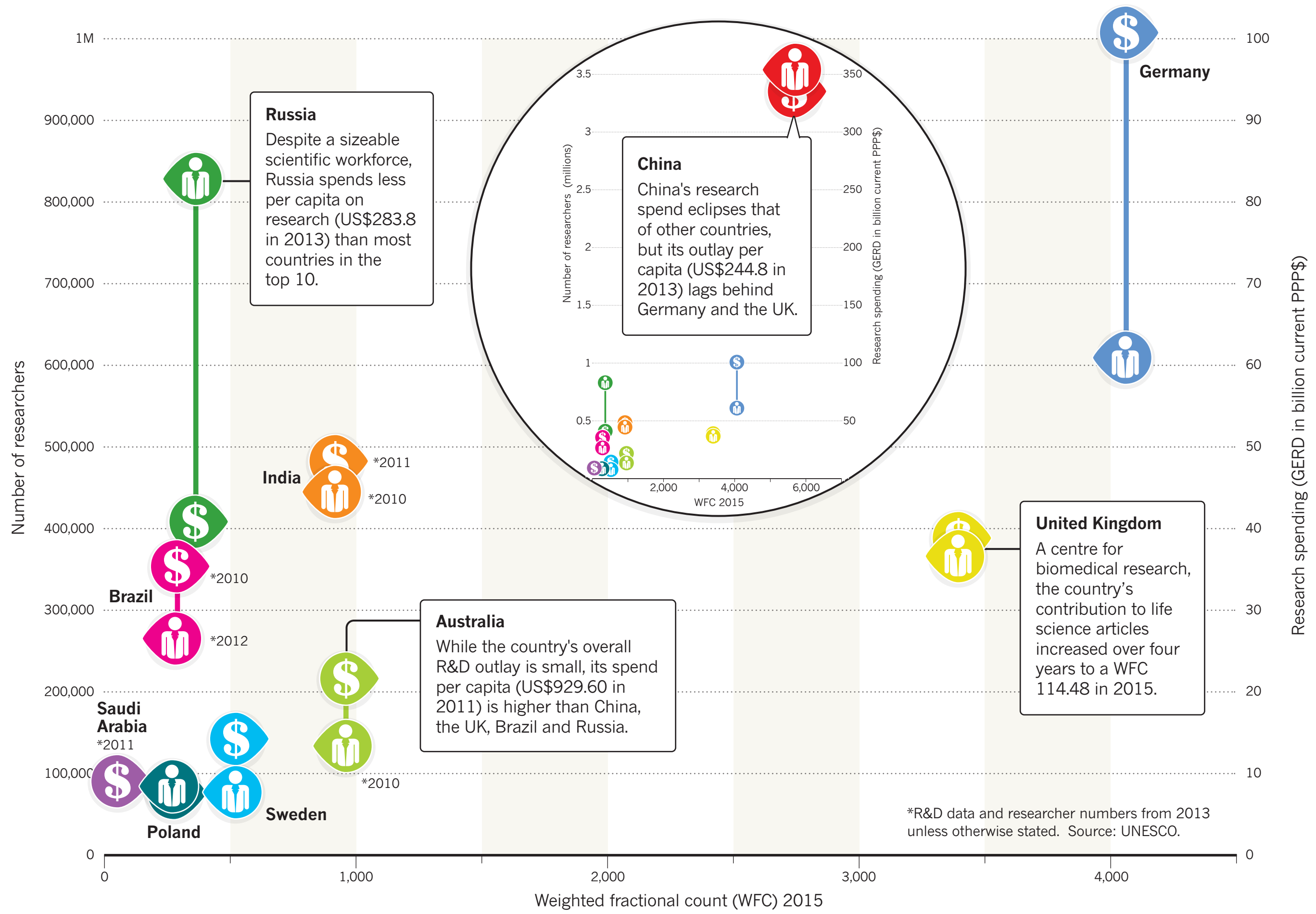
"Here there are still unsolved geological questions so Turkey is one of the hottest topics in world geoscience research. Every high-level Earth scientist wants to study here." **AM**

WHERE THE MONEY GOES

Resources for scientific research are spread unevenly throughout countries. Spending priorities vary according to budget constraints, historic research strengths and industry demands.

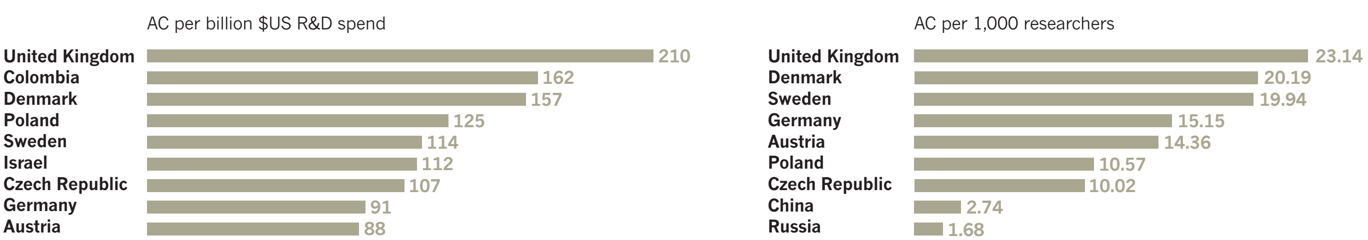
USING IT WISELY

Weighing a country's R&D spend against the size of its science workforce gives an indication of its research efficiency. Of the ten most improved countries in the index between 2012 and 2015, China has the biggest research budget and workforce, and contributes the most to the index. The United Kingdom is a strong performer, but spends less on a smaller workforce than China and Germany.



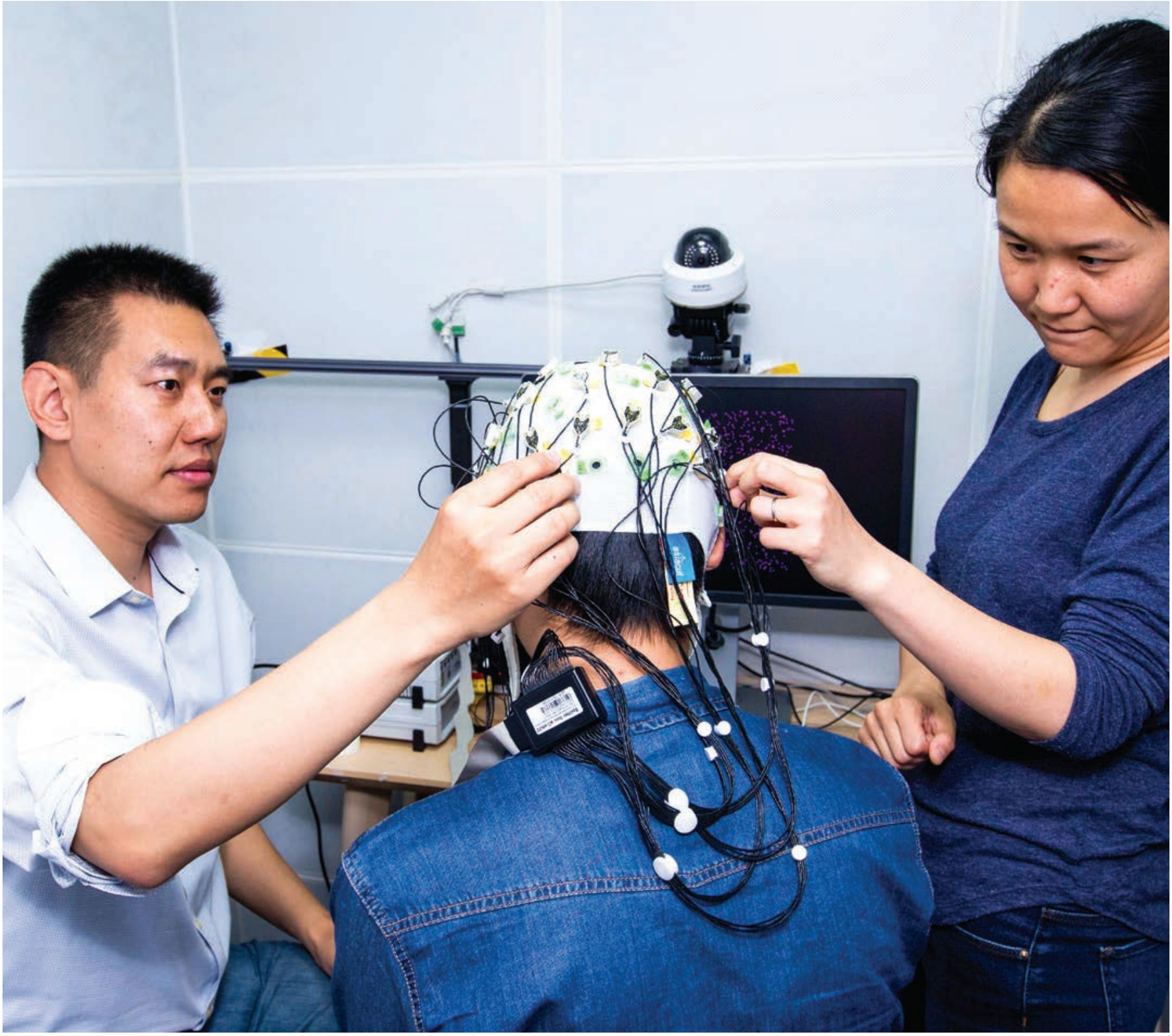
BANG FOR BUCK

When dividing a country's total number of articles in the index (AC) with their R&D spending and research workforce, the United Kingdom is the most efficient at maximizing its output per research dollar spent.



AC data are for 2015. GERD and researcher numbers data are from 2013. Countries with data available were included in analysis.

NYU SHANGHAI



Xing Tian and Qing Cai at the NYU-ECNU Institute of Brain and Cognitive Science study perception, action, language and their underlying mechanisms.

CATCH THEM IF YOU CAN

In institutions everywhere from super-cities to remote desert locations, researchers are setting the world of discovery on fire and pushing their institutions up the Nature Index rankings.

The performance of research strongholds like the Chinese Academy of Sciences and the University of Oxford have improved dramatically in the past few years.

As the research world's equivalent of Apple or Google, these well-established organizations have big budgets, some of the best brains and top-notch equipment.

So, what about the new kids on the block? As a database that tracks articles published in top-tier science journals, the Nature Index is well placed to identify the world's research rising stars.

The top 100 most improved institutions in the index between 2012 and 2015 are ranked by their contribution to 68 high-quality journals, a metric known as weighted fractional count (WFC). Here we profile 25 institutions. Some made the grade by improving most in their country, and others were highlighted for their meteoric rise up the index's global rankings.

China has so many institutions in the latter category, we could have produced a whole magazine about them.

SOOCHOW UNIVERSITY

CHINA

2012 WFC: 56.04

2015 WFC: 108.47

Founded in 1900 during the Qing dynasty, Soochow University is named after its birthplace of Suzhou on China's east coast — 30 minutes from Shanghai via bullet train. The institution has strengthened key areas of research over the past decade, a strategy that has culminated in the doubling of its Nature Index score since 2012. "The university decided to use its limited resources to strengthen a few selected departments," says Soochow materials science professor, Lee Shuit-Tong.

Research centres that focus on central government research priorities, such as nanotechnology, biomedicine and energy research, have boosted the university's published papers overall, while fuelling researchers' drive to produce higher quality research, says Lee.

"It was not until 2008 that the development of Soochow University drastically sped up," says Lee, the founder of the institute of functional nano and soft materials (FUNSOM), which opened in 2008 with US$22 million from the university. FUNSOM contributes more than 40% of the university's peer-reviewed articles published in high-impact journals, with only 2% of the total staff. Lee and a colleague were recognized by Thomson Reuters in its 'Highly Cited Researchers' list of 2015. Their team is exploring how cellular imaging can be used in cancer therapy and in the development of green-energy devices. **SO**

INSTITUTE FOR BASIC SCIENCE

SOUTH KOREA

2012 WFC: 1.04

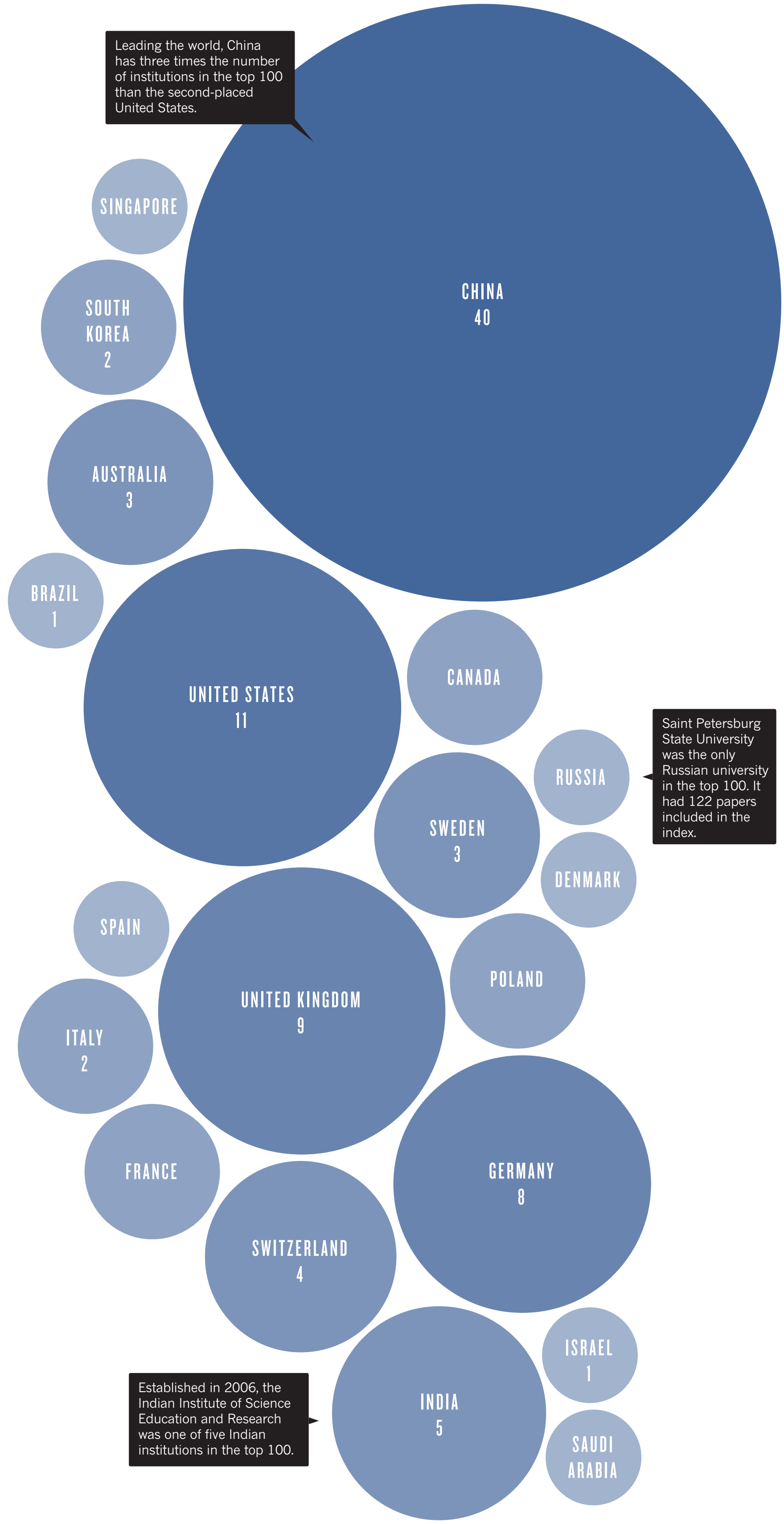
2015 WFC: 50.31

For decades, South Korea has propelled its economy by investing in applied research — often in tandem with industry — while basic science was neglected. The Institute for Basic Science (IBS) was intended to reverse that trend, becoming Korea's equivalent of Japan's RIKEN and Germany's Max Planck Institutes.

Founded in 2012, IBS is a new player even in a line-up of rising stars. It has grown to 26 centres with focuses ranging from dark matter to gene editing to graphene to nanomedicine. Each centre has an average annual budget of US$8.5 million, and there are plans to establish 50 centres

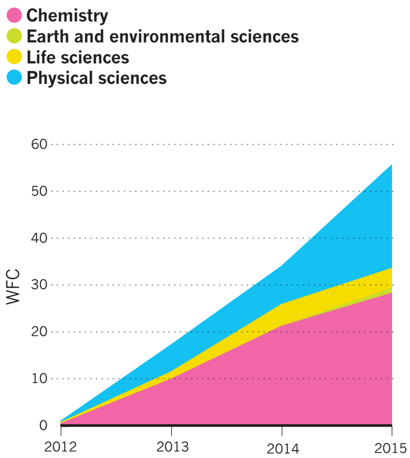
TOP 100 MOST IMPROVED

Countries are sized by the number of institutions in the top 100 institutions with the highest increase in their absolute contribution to the index, a metric known as weighted fractional count (WFC), between 2012 and 2015.



GROWTH SPURT

Since the Institute for Basic Science in South Korea was founded in 2012 its WFC has skyrocketed. In 2015, 189 papers were included in the index.



by 2021. About a third of these will be at a headquarters campus to be constructed in Daejeon; the rest hosted by universities and institutes around the country.

"The funding is generous," says Jin-soo Kim, who heads the Center for Genome Engineering. "I don't have to write a grant."

South Korean leaders hope IBS will deliver on the dream of a scientific Nobel prize. "The phenomenon is aggravated by Japanese success," says IBS president Doochul Kim, referring to the 21 scientific Nobels won by the country's rival. "Korean people have always compared and tried to catch up to Japan. Politicians ask us, 'What have you done?'"

South Korean president, Park Geun-hye, has also argued that IBS has many benefits. With Korea's economy slowing, the technological rise of China, and renewed competition from Japan spurred by a weak yen, she is betting on IBS and basic science as a catalyst for growth. **MZ**

EAST CHINA NORMAL UNIVERSITY

CHINA

2012 WFC: 35.55

2015 WFC: 83.90

East China Normal University (ECNU) evolved from a teacher training university to a multi-disciplinary institute and is fast gaining a reputation for its world-class research. Between 2012 and 2015, the university's contribution to articles in the index has more than doubled, lifting it more than 200 places in the Global 500 ranking to 142nd position in 2015.

The university prides itself on its international collaborations, and its most recent partnerships have led to fruitful research. In 2011, ECNU and New York University joined forces to establish China's first Sino-American university — New York University Shanghai (NYU Shanghai). A 2015 study by researchers at NYU Shanghai's Institute of Brain and Cognitive Science found physical evidence that the brain arranges words in a hierarchy using a form of 'internal grammar'. The findings, published in *Nature Neuroscience*, suggest the human mind can intuitively configure a string of words such as 'ancient history is drinking tea' into a structure that has meaning, such as 'drinking tea is ancient history'. **SO**

HUNAN UNIVERSITY

CHINA

2012 WFC: 49.22

2015 WFC: 92.47

Hunan University is one of China's oldest and can trace its roots back to the year 976 CE. The campus sits on the Xiang River in Changsha, the capital city of the province of Hunan.

The university has almost doubled its contribution to the index between 2012 and 2015, a rise attributed to its recruitment of top-level researchers, according to Weihong Tan, a professor of chemistry and biomedical engineering. In 2013, Hunan established the Institute of Chemical Biology and Nanomedicine under the direction of leading US scientists, Chad Mirkin from Northwestern University and David Walt from Tufts University.

Tan's team are focusing on research in biomedicine, which he hopes will lead to new industries for the province. **SO**

SICHUAN UNIVERSITY

CHINA

2012 WFC: 44.88

2015 WFC: 83.22

Sichuan University's position in the Nature Index has soared in recent years, climbing from 278th place in the Global 500 institutions ranking in 2012 to 145th in 2015. New facilities, well-funded salary start-up packages and accessible research grants have attracted eager young researchers to the university in Chengdu in southwest China. Fifty international researchers have been recruited in the past two years to the university's National Key Laboratory of Biotherapy, which opened in 2005. The multidisciplinary research centre enables discovery and development of innovative drug candidates.

"Basic research, preclinical development, translational and clinical medicine are seamlessly integrated," says Wei Yuquan, who is director of the laboratory as well as the university's vice president.

He says: "We want to bring disciplines together to try to solve the major problems in medicine, science and technology." **SO**

KING ABDULLAH UNIVERSITY OF SCIENCE & TECHNOLOGY

SAUDI ARABIA

2012 WFC: 40.51

2015 WFC: 72.19

KAUST is more than just a university — it's practically a city. Occupying 36 square kilometres on the Red Sea, KAUST is separated from conservative Saudi society. Courses are coeducational, women can drive and are not required to wear veils, and the campus has one of only two sanctioned movie theatres in the kingdom.

The graduate-only research university was founded in 2009 at the behest of the late King Abdullah. He hoped to create a premier institute to kickstart Saudi Arabia's research scene — and to help develop clean energy technology to wean it off oil. He left the university with an endowment estimated at US$20 billion, and it recruited foreign researchers with start-up funds, reportedly worth up to US$1 million each, to set up labs. In the early days many faculty members left, claiming the university had failed to deliver on its promises. But the institution's contribution to high-quality research had surged by 2015 to levels 80% higher than in 2012. **MZ**

> ***"We want to bring disciplines together to solve major problems in science and medicine."***

NANYANG TECHNOLOGICAL UNIVERSITY

SINGAPORE

2012 WFC: 177.60

2015 WFC: 207.83

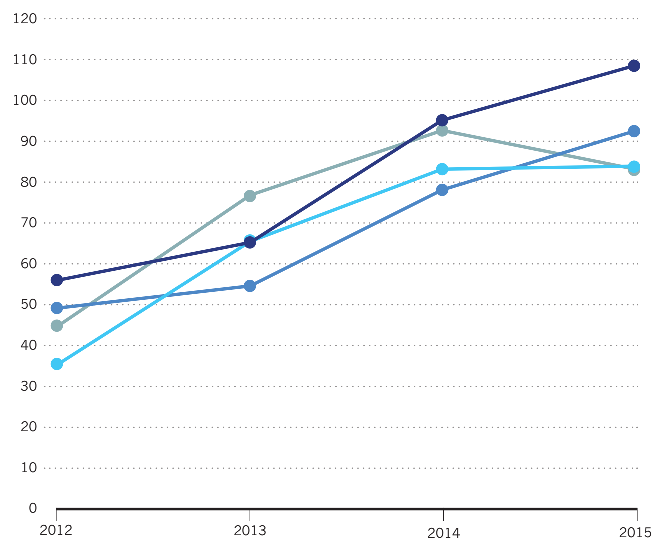
Nanyang Technological University is one of Asia's research powerhouses, ranked 37th in the index's Global 500. Most of the Singaporean university's high-quality research output is in chemistry, but it has

ANASTASIA KHRENOVA/KAUST

KAUST's Center for Desert Agriculture works toward food security solutions in arid climates.

CHINA'S DOMINATION

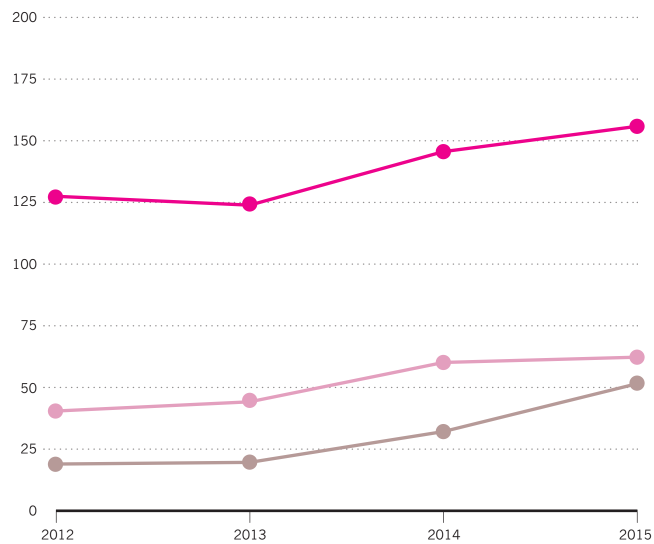
The upward trajectory of many of China's less well-known research universities is reflected in the WFC growth of four Chinese universities in the top 100.



- Soochow University
- Hunan University (HNU)
- East China Normal University (ECNU)
- Sichuan University (SCU)

CHEMISTRY CHAMPS

Chemistry output was a significant driver of the performance for several research institutions in the top 100 most improved institutions in the index.



- Nanyang Technological University (NTU)
- Indian Institute of Science Education & Research (IISER)
- King Abdullah University of Science & Technology (KAUST)

also invested heavily in sustainability and environmental research.

Founded in 1991 through a merger between an engineering university and the national institute for teacher training, it was granted autonomy from the state in 2006 and swiftly transformed into a research-intensive institute, taking advantage of increased public funds. **MZ**

INDIAN INSTITUTE OF SCIENCE EDUCATION & RESEARCH

INDIA

2012 WFC: 50.97

2015 WFC: 78.67

The Indian Institute of Science Education and Research (IISER), in Pune, was only founded in 2006. A decade later, the Indian government has established six additional IISERs, in Kolkata, Mohali, Bhopal, Thiruvanthapuram, Tirupati and, most recently, in Berhampur. Collectively, they have risen to become the fourth-ranked Indian organization in the Global 500 by boosting their contribution to high-quality science by more than 50% since 2012.

The IISERs were set up to build on the success of India's highest entrant to the Global 500, the Indian Institutes of Technology, says Haripriya Rangan from the Australia India Institute in Melbourne. "We have a very strong reputation overseas, but in India there has been not much reflection of that," she says.

The IISERs, with an emphasis on fundamental scientific research, might address that. "They have been well funded, with state-of-the-art facilities," says Kartik Shanker, an ecologist at the IIT in Bangalore. "They have hired good young, faculty with very active research programmes."

But, the IISERs have their critics. The institutions were set up to span the fundamental sciences, and yet there has been a perceived bias toward chemistry, sources say. The Scientific Advisory Council to the Prime Minister, which recommended the IISERs be established, is headed by a prominent chemist, and chemists were appointed as directors to four of the first five IISERs to open. In the index, chemistry research accounts for almost 80% of the organizations' contribution to world-class natural science research. **JMC**

> ***"We have a strong reputation overseas, but in India there is not much reflection of that."***

ISRAEL INSTITUTE OF TECHNOLOGY (THE TECHNION)

ISRAEL

2012 WFC: 68.75

2015 WFC: 95.59

It's difficult to pinpoint the factors that have pushed the Technion up the Nature Index Global 500 ranking of top-performing research organisations, according to Ehud Keinan, a chemist at the institution and, since 2009, the president of the Israel Chemical Society. "If the recipe was easy and clear, everybody would do it." But the Technion's chemists have certainly got some of the reagents right.

Established in the port city of Haifa in 1912, Israel's oldest university has 18 academic departments, including all scientific areas from engineering to medicine. But, of the university's 565 academic faculty, it's primarily the chemists that have powered the Technion's rise up the index rankings, from position 179 in 2012 to 108 in 2015. They have more than doubled their contribution to chemistry papers published in top-tier journals since 2012, the first year the index was compiled.

The Technion's overall contribution to high-quality science papers jumped almost 40% in that time.

Keinan identifies one probable reason for the institute's performance: the awarding of the 2004 Nobel Prize in chemistry to Technion researchers Aaron Ciechanover and Avram Hershko for their discovery of how the body tags proteins for recycling. This success was followed in 2011 by Technion chemist Dan Shechtman who won the same gong for his discovery of quasicrystals, an unexpected form of matter that combines order and disorder.

"Success is an autocatalytic process," Keinan says. "When you get good faculty, you can attract even better faculty members, and good graduate students."

Keinan is now working to establish a branch of the Technion in Guangdong, China, to leverage its strength in chemistry. **JMC**

BAYLOR COLLEGE OF MEDICINE

The neurosurgery department at Baylor College of Medicine, whose other areas of strength are paediatrics, immunology and the study of cardiovascular disease.

BAYLOR COLLEGE OF MEDICINE

UNITED STATES

2012 WFC: 66.29

2015 WFC: 92.15

Baylor College of Medicine (BCM) is the self-proclaimed 'intellectual heart' of the world's largest medical complex, Houston's Texas Medical Centre.

Despite its enormity now, BCM was small for much of its 116-year history. A key figure in setting the college on its current trajectory was a paediatrician called Ralph Feigin.

"Feigin was an impressive dude and a very successful leader," says Scott Simon, a former faculty member, now at the University of California, Davis. Feigin led the paediatrics department from 1977 to 2008, and was Baylor's president and CEO from 1996 to 2003. "I've never seen a physician who was as good an administrator," Simon adds.

Baylor has grown by building on existing strengths, says Simon — including paediatrics, immunology and cardiovascular disease. "They went for three or four areas they did really well at, and pulled in the best people — and then other programmes grew around it."

In 2009 and 2010, BCM considered merging with neighbouring Rice University or with Baylor University in Dallas, from which it had split in 1969 — but decided to stay solo. Remaining independent hasn't hurt its output: BCM's contribution to high-quality research rose 39% from 2012 to 2015. **JMC, MZ**

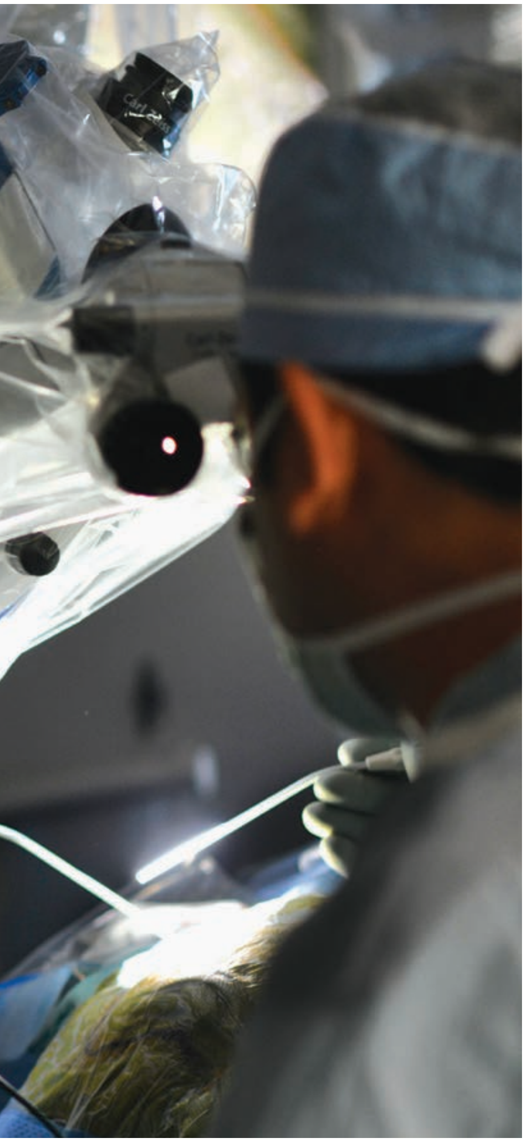
NORTH CAROLINA STATE UNIVERSITY

UNITED STATES

2012 WFC: 51.71

2015 WFC: 76.29

Located in Raleigh, North Carolina State University is at the heart of a region quickly becoming a major research and technology hub. In 2015, *Forbes* magazine reported it had the highest rate of STEM job growth in the US for the preceding decade, topping



Silicon Valley. NCSU is one of the three academic anchors of the so-called Research Triangle, along with Duke University in Durham and the University of North Carolina in Chapel Hill. More than 100 start-ups or spinoffs have emerged from North Carolina State, attracting $US1.6 billion in venture capital.

Both its total number of articles in the index and its contribution to those articles spiked by roughly 50% between 2012 and 2015, vaulting 91 places in the index's Global 500 to rank 164. Traditionally strong in engineering, it hosts two United States National Science Foundation (NSF) Engineering Research Centers: one that focuses on smart electrical grids and one specializing on health wearables.

In the mid-2000s, North Carolina State become mired in controversy by claims that three of its scientists had falsified data in a 2004 study finding that RNA could regulate the growth of the metal palladium. A university investigation in 2008 found the paper, which had been published in *Science*, included false data, but stopped short of finding intentional misconduct. A more thorough 2013 investigation by the NSF inspector general concluded the pair had indeed "recklessly falsified data," although the NSF eventually ruled it had not been intentional. *Science* retracted the paper in February 2016. The scientists in question no longer work at the university. **MZ**

XI'AN JIAOTONG UNIVERSITY

CHINA

2012 WFC: 30.27

2015 WFC: 50.78

In the past four years, Xi'an Jiaotong University, in the ancient former capital city of Xi'an in central China, increased its high-quality natural science in all subject areas, but most dramatically in chemistry and the physical sciences.

The university is home to the State Key Laboratory for Mechanical Behavior of Materials. State key laboratories are centrally-funded institutions that are designed to funnel resources to government priority areas. Materials science and its industrial applications are considered key to China's economic development.

The team at Xi'an Jiaotong are exploring ways in which materials behave at the nano scale. They recently developed a nanospring that can efficiently store and release mechanical energy, a component that could be useful in more complex nanomechanical systems.

Four years ago, Xi'an Jiaotong established a multidisciplinary research institute called the Frontier Institute of Science and Technology (FIST) which now has 11 centres focusing on life sciences, medicine and mathematics, and has recruited 40 international scientists. Among them are Timothy Bliss and Graham Collingridge from the UK, who in March 2016 were awarded the European 'brain prize' for their work on memory mechanisms. **SO**

HARBIN INSTITUTE OF TECHNOLOGY

CHINA

2012 WFC: 19.30

2015 WFC: 39.73

In the capital of the far northern province of Heilongjiang, the Harbin Institute of Technology (HIT) is synonymous with research supporting the country's space science programme. China's goals include establishing a space station, sending astronauts to Mars, and industrializing rocket technology to more efficiently transport vessels into space.

"China's goals include establishing a space station and sending astronauts to Mars."

Over the past decade, Harbin's researchers have worked on China's Shenzhou spacecraft programme, including its first manned spacecraft, Shenzhou 5, and more recently, a satellite for emergency data monitoring and imaging, Kuaizhou-1, which was launched in 2013.

The university, established in 1920, is a member of China's elite C9 alliance — nine institutes in China that receive 10% of the country's research and development funding, and are relied on by the central government to produce world-class results.

HIT has experienced rapid growth in the index, doubling its contribution to top journals in four years and moving up 208 places in the Global 500 to reach 304th position in 2015. **SO**

UNIVERSITY OF REGENSBURG

GERMANY

2012 WFC: 56.86

2015 WFC: 72.58

The University of Regensburg's improvement in the index is driven mostly by chemistry, with the university placing 11th among the country's academic chemistry institutions in 2015.

As head of the chemistry department, Oliver Reiser points out, "that puts Regensburg above many universities a lot bigger than us." The department has a mere 18 tenured academics, compared to more than 25 at the University of Stuttgart (ranked 13th) and more than 40 at the Ruhi University Bochum (19th).

In 2010, Regensburg, in south-east Germany, won special status from the national funding agency for its studies of photocatalysis — chemical reactions accelerated by sustainable sunlight — which secured the department additional funding until 2019.

Regensburg's physics department hosts a similar centre researching carbon-based nanostructures. Overall, the university's contribution to index papers grew by almost 28%, the largest surge by any German institution in the top 100 most improved institutions. **JMC**

A. ABRUSCI/ISTITUTO ITALIANO DI TECNOLOGIA

ITALIAN INSTITUTE OF TECHNOLOGY
ITALY

2012 WFC: 22.42

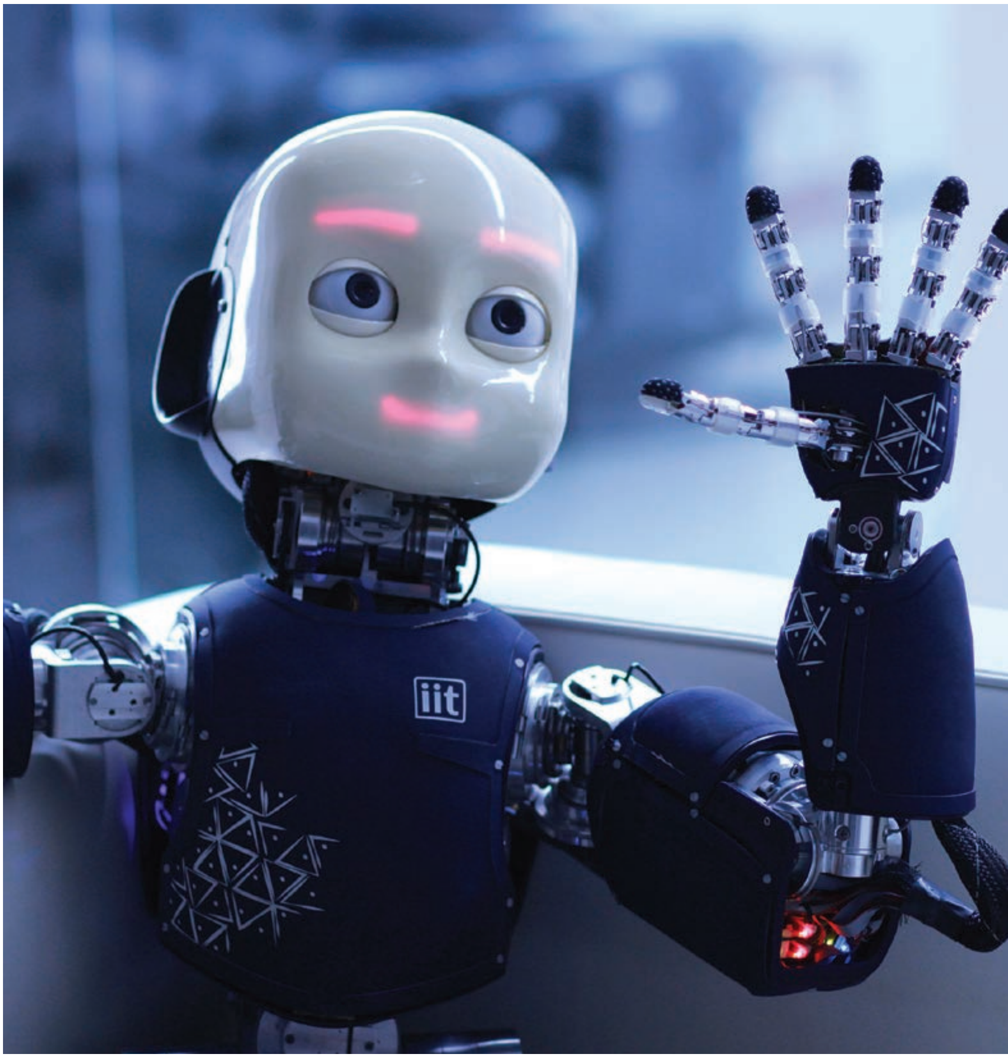
2015 WFC: 41.39

When the Italian government decided to establish a research organization in 2004, it sought inspiration from the world-renowned research organizations of Germany, in particular the Max Planck Society.

The Max Planck Society is funded by the state, but operates as a private foundation. "They are a little bit more agile than universities," says Roberto Cingolani, scientific director at the Italian Institute of Technology, which opened its main laboratory in Genoa in 2009.

The institute's 1,300 research staff are spread across 11 labs around the country and two small units at premier US universities, Harvard and the Massachusetts Institute of Technology. Since 2012, the institute's scientists increased their contribution to index papers more than any other Italian institution.

The institute concentrates on nanotechnology, robotics and brain science. The work is more connected than it might first sound. "Our general mission is 'translating evolution into technology'," Cingolani says. The institute's best-known piece of research is its humanoid research robot, iCub, developed to test the idea that a robot, like an infant human, can develop an understanding of the world around it via physical interaction with its environment. **JMC**



The Italian Institute of Technology's iCub was built to test whether a robot can learn through physical interaction.

UNIVERSITY OF SUSSEX
UNITED KINGDOM

2012 WFC: 18.09

2015 WFC: 33.26

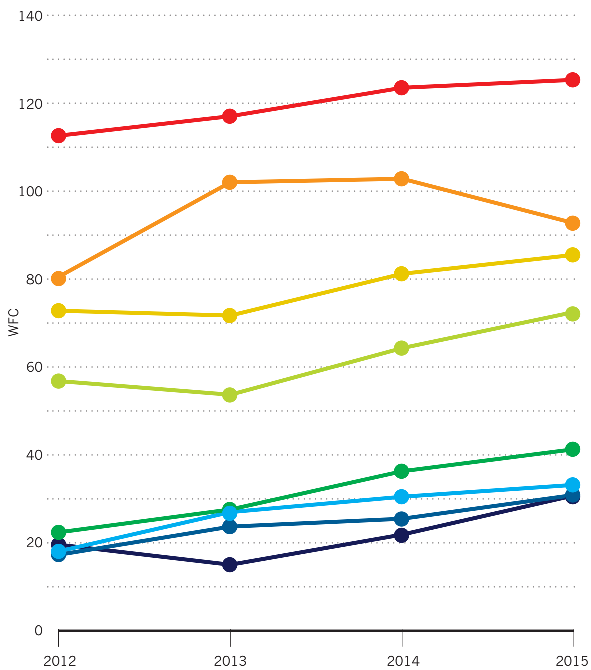
There's no great mystery behind the rapid rise of the University of Sussex in the index. In 2012, the university didn't place in the first Global 500 ranking. But that year, it began strategic financial investment in key disciplines, in particular, physics and astronomy. The number of research staff has increased by more than 60%, from 23 in 2012/13 to 40 today and is still growing, says astrophysicist, Peter Coles, who leads the school of mathematical and physical sciences.

Recruits were brought in to reinforce existing strengths in astronomy and particle physics, and to establish programmes in quantum technologies and condensed matter physics.

Mainly thanks to a quadrupling of its physical sciences score, the university

EUROPE'S RISING STARS

As a continent, Europe is home to the most rising stars in the index. On the right is a timeline of eight institutions by WFC between 2012 and 2015.



reached 351 in the Global 500 in 2015. That represents an 83.9% rise in its contribution to index papers since 2012 — the biggest jump of any UK research organisation in the top 100 most improved institutions. **JMC**

PERIMETER INSTITUTE FOR THEORETICAL PHYSICS

CANADA

2012 WFC: 32.26

2015 WFC: 46.54

When Canadian prime minister, Justin Trudeau, visited the Perimeter Institute for Theoretical Physics in Waterloo, Ontario in April 2016, he wowed reporters by nailing a 30-second explanation of quantum computing that instantly went viral. But, the telegenic PM's visit didn't just burnish his image. He announced the Ontario government would contribute CDN$50 million (US$38 million) to its budget over five years.

That's only one funding avenue this unique public–private institute attracts. Since its founding in 2000, it has lured funding from foundations, corporate sponsors, and philanthropists — including CND$150 million from its founder, Mike Lazaridis, also the co-founder of BlackBerry. The institute now has 25 full-time faculty and counts famed theoretical physicists Stephen Hawking and Leonard Susskind among its 160 scientists-in-residence. The institute's contribution to index papers increased by 44% between 2012 and 2015, pushing it from position 362 in the index's Global 500 ranking in 2012 to position 267 in 2015. **MZ**

JAGIELLONIAN UNIVERSITY

POLAND

2012 WFC: 17.37

2015 WFC: 31

The standing of Jagiellonian University in Krakow, Poland, has fluctuated since its foundation in 1364, often buffeted by shifts in the balance of power in central Europe. But, the oldest university in Poland is also the country's rising star, having increased its contribution to high-quality research papers by 79%, the largest increase of any Polish institution in the top 100 most improved institutions.

In 2010, Jagiellonian was awarded a European Union regional development grant to host Poland's first synchrotron. Completed in late 2015, the synchrotron has yet to contribute directly to the university's performance in the index. However, the synchrotron was integral to Jagiellonian's successful 2012 bid, with partner research organizations in Krakow, to become Poland's flagship research centre for physics and chemistry — a status that guarantees additional government funding. The research financed by this grant is already helping boost Jagiellonian up the Global 500 in which it ranked 370 in 2015. **JMC**

FRENCH NATIONAL INSTITUTE FOR HEALTH AND MEDICAL RESEARCH

FRANCE

2012 WFC: 72.95

2015 WFC: 85.55

The French government shook up its science funding in 2009, introducing a €47 billion 'Investments for the Future' fund to boost national competitiveness in targeted sectors. Some €5 billion of the fund was allocated to biomedical research, according to Yves Lévy, CEO of the French National Institute for Health and Medical Research (INSERM). "Given INSERM's focus, we were very well placed and have won many of these grants," Lévy says. All INSERM's researchers work in research units embedded in French universities, most of them in university hospitals.

Since 2012, INSERMS's contribution to the index grew the most of any French institution in the top 100 most improved list, and it placed 136 in the Global 500 ranking in 2015. But as Lévy notes, some of its highest impact research is published in medical journals such as the *New England Journal of Medicine*, which are not included in the index. **JMC**

UNIVERSITY OF COPENHAGEN

DENMARK

2012 WFC: 112.63

2015 WFC: 125.32

The oldest and most prestigious university in Denmark, the University of Copenhagen ranked 72nd globally in the index in 2015, with strengths in life and physical sciences. The university's researchers published 25% more articles in top natural science journals in the index in 2015 than in 2012.

The rise came despite declines in state funding, which the university sought to offset with external sources, including six major laureate grants from the Novo Nordisk Foundation between 2012 and 2014 totalling US$36 million.

But, the funding cuts are now a major crisis that could jeopardize the university's gains: In February 2016, due to budget constraints, the university announced it would lay off 7% of staff — 500 people, including more than 100 researchers — and impose a hiring freeze. **MZ**

SÃO PAULO STATE UNIVERSITY

BRAZIL

2012 WFC: 11.21

2015 WFC: 23.52

São Paulo State University (UNESP) is fairly new to Brazil's burgeoning research scene, but it arrived with intent. Founded in 1976 by a merger of previously independent institutions, this multi-campus university is modelled on the University of California's state-wide system and consists of 24 campuses across São Paulo State, including its eponymous capital. In 2009, it embarked on a hiring frenzy of more than 1,000 faculty and researchers, and from 2012 to 2015, its contribution to high-quality science in the index more than doubled.

The university's research strengths are in physics and astronomy. Its Institute for Theoretical Physics was founded in 1951, predating the university itself. It also hosts the ICTP South American Institute for Fundamental Research, an offshoot of the International Centre for Theoretical Physics in Trieste, Italy.

However, Brazil's economic crisis has put science funding on the chopping block, cutting it by 25% in 2015. After the May impeachment of President Dilma Rousseff, the science ministry was merged with the telecommunications ministry, leaving many researchers wondering if science will remain downgraded for many years to come. **MZ**

UNIVERSITY OF BASEL

SWITZERLAND

2012 WFC: 80.50

2015 WFC: 92.77

Switzerland's city of Basel is a hub for life sciences research and home to the headquarters of pharmaceutical giants, Novartis and Roche. In 2007, the University of Basel decided to play to this strength and make life science its key focus.

This concentration of effort has helped power the university's 40-place leap up the Global 500 since 2012, to 109 in 2015.

The university's strategy was supported by the regional government, says Christian

CHRIS SPENCER/TRAVELINGGEOLOGIST

Chris Spencer of Curtin University at the basalt outcrop of the Semail Ophiolite in Wadi Jizzi, Oman.

Sengstag, deputy vice-rector for research at the university. “Lots of resources have been pumped into life science research,” he says. New research groups have been established, and the university has invested “in top-notch infrastructure and buildings.” Last year Basel scientists used single-molecule spectroscopy to show how E. *coli* bacteria fold the proteins that form their outer membrane — a long-standing drug discovery target.

Political support can be fickle, however. Federal elections held in 2015 saw a strong swing toward the right-wing SVP party, which has vowed to cut government spending. “In the future we will probably have to live on a lower budget,” says Sengstag. **JMC**

SAINT PETERSBURG STATE UNIVERSITY

RUSSIA

2012 WFC: 13.97

2015 WFC: 25.36

Saint Petersburg State University is known for educating generations of Russia’s elite: it counts current president Vladimir Putin and prime minister Dmitry Medvedev among its alumni.

After the fall of the USSR in 1991, funding plummeted and national science output slipped. But, budgets are slowly building up, and the country’s oldest university leapt into the Global 500 in the index in 2015, at position 410, by more than doubling its contribution to high-quality science over the previous four years.

Part of its upswing has come from increased state funds, including seven projects funded by the Russian education and science ministry’s so-called ‘mega-grants’. These grants of 90–150 million roubles (US $1.3–2.2 million each) are open to scientists around the globe, and have played a key role in the nation’s strategy to attract foreign talent and revitalize its research scene. Three of the university’s mega-grant projects involve nanomaterials, reflecting a traditional strength in chemistry and physical sciences, but others focus on the geomorphology of the polar regions, genomics and bioinformatics. **MZ**

UMEÅ UNIVERSITY

SWEDEN

2012 WFC: 19.64

2015 WFC: 30.84

Sweden’s temperate environment is a world away from tropical Latin America, where the Zika virus continues to wreak havoc, but researchers at the country’s Umeå University have joined the fight to stop Zika spreading further. Global health is one of Umea’s key strengths, and that includes an EU-funded consortium, established at the university in 2011 to fight Dengue fever. (Zika and Dengue are transmitted by some of the same species of mosquito.)

Umeå University’s contribution to high-quality science increased by 57% from 2012 to 2015, placing it in the top 400 institutions globally. Founded in 1965, the university prioritizes biotechnology and life sciences, and its regional forests in particular have yielded plenty of research. In 2013, The Umeå Plant Science Center, a collaboration between Umeå University and the Swedish University of Agricultural Sciences, mapped the Norway spruce genome, which comprises 20 billion base pairs — more than six times as many as the human genome. Its energy research portfolio also includes developing biofuels from trees. **MZ**

CURTIN UNIVERSITY

AUSTRALIA

2012 WFC: 13.23

2015 WFC: 24.16

The state of Western Australia has two distinguishing qualities: it is rich in mineral resources and its interior is sparsely populated. Both factors feed into Curtin University’s focus on Earth sciences.

A decade ago, the university, located in Perth, reorganized its mining school to incorporate key aspects of applied geology, says Graeme Wright, Curtin’s deputy vice-chancellor for research. Wright cites this transformation as the key to the 54% increase in Curtin’s contribution to articles in Earth and environmental sciences in the index since 2012. Overall, the university’s contribution to journals in the index increased more than 80%.

Western Australia’s vast, people-free interior also makes it among the most radio-quiet places on the planet and an ideal site for the Square Kilometre Array (SKA), an internationally funded next-generation radio telescope, scheduled to be fully operational in 2030. The SKA will peer deeper into the universe than ever before and help answer questions about how stars, galaxies and the universe formed and evolved. The state’s ‘pathfinder’ radio telescopes are already boosting Curtin’s physical sciences index score. “From small beginnings, we built a strong radioastronomy group,” says Wright.

For Wright, the key to Curtin’s index success is its collaborations. “We are a modestly-sized university. We know we’ve got great people, but that there are even better people outside. That’s key to what we do.” **JMC**

A GUIDE TO THE NATURE INDEX

A description of the terminology and methodology used in this supplement, and a guide to the functionality available free online at natureindex.com.

The Nature Index is a database of author affiliations and institutional relationships. The index tracks contributions to articles published in a group of highly selective science journals, chosen by an independent group of active researchers.

The Nature Index provides absolute counts of publication productivity at the institutional and national level and, as such, is one indicator of global high-quality research output.

Data in the Nature Index are updated monthly, with the most recent 12 months of data made available under a Creative Commons licence at natureindex.com.

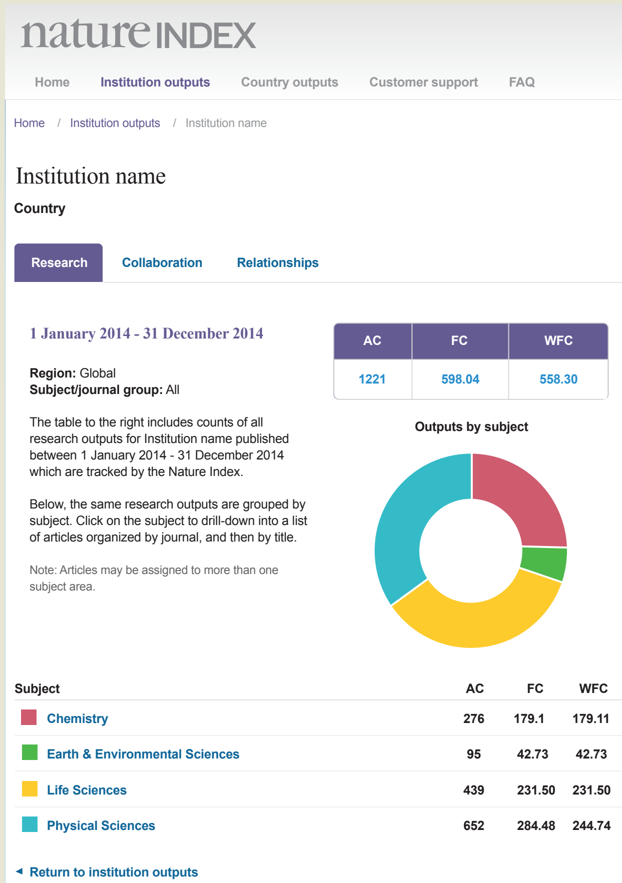
The database is compiled by Nature Publishing Group (NPG). The list of journals tracked by the Nature Index is under review, and from 2016 will be extended to include the clinical sciences.

NATUREINDEX.COM

A global indicator of high-quality research

natureindex.com users can search for specific institutions or countries and generate their own reports, ordered by article count (AC), fractional count (FC) or weighted fractional count (WFC).

Each query will return a profile page that lists the country or institution's recent research outputs, from which it is possible to drill down for more information. For example, articles can be displayed by journal, and then by article title. As in the supplement, research outputs are organized by subject area. The profile page also lists the institution or country's top collaborators, as well as its relationship with other research organizations.



NATURE INDEX METRICS

There are four measures provided by the Nature Index to track affiliation data. The simplest is the *article count (AC)*. A country or institution is given an AC of 1 for each article that has at least one author from that country or institution. This is the case whether an article has one or a hundred authors, and it means that the same article can contribute to the AC of multiple countries or institutions.

To get a sense of a country or institution's contribution to an article, and to remove the possibility of counting articles more than once, the Nature Index uses the *fractional count* (FC), which takes into account the relative contribution of each author to an article. The total FC available per paper is 1, which is shared between all authors under the assumption that each contributed equally. For instance, a paper with 10 authors means that each author receives an FC of 0.1. For authors who have joint affiliations, the individual FC is then split equally between each affiliation.

The third measure used is the *weighted fractional count (WFC)*, which applies a weighting to the FC to adjust for the over-representation of papers in astronomy and astrophysics. The four journals in these disciplines publish about 50% of all papers in international journals in this field — approximately five times the equivalent percentage for other fields. Therefore, although the data for astronomy and astrophysics are compiled in the same way as for all other disciplines, articles from these journals are assigned one-fifth the weight of other articles (i.e., the FC is multiplied by 0.2 to derive the WFC).

The total FC or WFC for an institution is calculated by summing the FC or WFC for individual authors.

The process is similar for countries, although complicated by the fact that some institutions have overseas labs that will be counted towards their host country totals. What's more, there is great variability in the way authors present their affiliations.

Every effort is made to count affiliations consistently, with a background of reasonable assumptions.

The fourth metric is *collaboration score*. This is a relatively new metric that is derived by taking the average of the FCs for all the bilateral relationships for that institution or country.

If institution A has relationships with two others, B and C, then the collaboration score is average of the FC for A + B and A + C.

For more information on how the affiliation information is processed and counted, please see the FAQ section at natureindex.com.

THE SUPPLEMENT

Nature Index 2016 Rising Stars is based on data from the Nature Index, covering articles published during four consecutive years from 1 January 2012 to 31 December 2015.

Most analyses within the supplement use WFC as the primary metric, as it provides a more even basis for comparison across multiple disciplines, and in determining the relative contribution of each country or institution. ■